

# **COVID-19 Workbench**

### **Features & Benefits**

**March 2021** 

Provided in partnership with **Aridhia** Informatics



#### What is the COVID-19 Workbench?





### **The Workbench for Data Contributors**

Contribute data to pandemic-related research in a controlled environment.

#### Control

- Access and authorisation approvals to your dataset
- Research proposals using your data

#### Services

- Templated data sharing agreements
- Researcher accreditation and project
   approval
- Metadata sharing and publishing
- Tiered, federated data sharing protocol
- Managed data access requests

#### Compliance

- Certified security and information governance
- Choose the location of your data





Audit, reporting & governance

### **The Workbench for Researchers**

Access valuable data in a trusted research environment with the opportunity to work on important research questions.

### 1

#### Data Access

- Streamlined accreditation
- Data discovery, data access requests

#### Tooling

- Collaborative workspaces with open source tooling (R, Python, SQL)
- Bring your own data, code and 3<sup>rd</sup> party software
- Specialised clinical research tools from ICODA partners

#### Compute

- Web based tools or clustered virtual machines (Windows, Linux, GPU)
- Scalable cloud computing





#### **Information governance**





### How do we ensure robust Data governance?

#### Safe People – Login & Access

- Accreditation process
- Invited users self-register
- ✓ **Login**: via 2 Factor Authentication
- International access: allowed

#### Safe Projects

 All project proposals reviewed for scientific merit, patient benefit, appropriate statistical analysis plans, facilitation of high quality research, availability of relevant data

#### Safe Data – Data Access Mechanisms

- Data provisioning: Performed by authorised Workspace administrators
- Receive data: Both UI and API file upload goes through malware scan and approval process
- ✓ Record linkage:
  - Linked data: yes, approved datasets only
  - Sensitive data: yes, approved datasets only. De-identification
  - Open data: yes, approved datasets only

#### Safe Output – Data Output/export

- All data/results export goes through a secure Airlock process
- Output checking and Statistical Disclosure Control process



### **Tools and Compute Services available**

#### **Compute & Services**

- Single-tenanted managed platform running on Microsoft Azure cloud
- Support for a wide range of compute options
  - ✓ Windows, Linux, general purpose, compute optimised, memory optimised, GPU
- Integration with Azure services, e.g. Azure ML
- Managed Data Analytics
  - R, Shiny, SQL, Python, Jupyter, Statistical Analytics Modules
- Federated Queries: Supported with selected data repository partners
- Federated Analytics: Supported with selected data repository partners

#### Software package access

- ✓ Default software: R, Shiny, SQL, Statistical Analytics Modules, Libre Office, Python, Jupyter
- ✓ Third party tooling: Meta analysis, Visual Analytics tools from Cytel, CODEx from Certara
- ✓ Software installation: as per an approved whitelist (allow list).
- ✓ Further Software/Tooling additions may be possible on agreement with Data Steward/Administrator
- ✓ Code/library import: as per an approved whitelist.
- Collaboration: Gitlab integration available





### Workspaces – built in features (Screenshots)









Host applications



### **Example tools: Visualization App**

- The visualization app is designed to provide analysts with an easy to use and a powerful tool to describe and visualize trials
- It can work as a collaboration tool for real time data exploration
- Features include
  - scatter plots, dot plots, boxplots, bar plots, histograms, densities and summary statistics tables
  - data manipulations such as categorize/cut, merge factor levels, recode/reorder categories, combine variables
  - summary/regression functions such as Smooth/Linear/Logistic Regressions, Mean Confidence Intervals, Median Prediction Intervals, Kaplan-Meier curves, Correlation Coefficient



C/Ubers/smoukses/Deskt	og/"Diny/projects/packages/gppsickeda/	inst/shinyapp - Shiny			- 0 X
Alaama to	agguickedel 0 f	1.0			"S Reputitor
velcome to	o ggquickeda! U.	1.0			
1		X/Y Plot Export Plots	Experimental Pioty 1	Descriptive Stats Data	Plot Code
Inputs Graph	Options How Ta	All Pabents	Al Patients		1-0-1
X <sup>1</sup> Y Axes LogLabers Graph Size/Zoom			North America	-	
		Geographic Region	Europe-		
The State of the S			White-		
Reference Lines/Target		Race	Black-		H
Themes and Manual Scales Options			Other +		
			Hisperic or Latino		1
Y axis label		-	Not Hispanic or Latine -	-	
		Age	18 to < 40 yr + 40 to < 65 yr -		
X axis label			>+ 65 jr		•
Recovery Rate Ratio (95% Cr/inPlacebo Better		Sex	Male-		
Plot Title			er 10 days-		1.0.1
			> 10 days -		1 • · · ·
Plot Subtitle		Summer and the summer	4 (not receiving oxygen) -		
		Baseline Ordinal Score	6 (receiving high-flow axygen) -		•
Title Positioning.			7 (receicing mi or ECMO)		4
Plot Caption				Recovery F	Rate Ratio (95% CI)
		Placebo Better Remdesivir Better			
Caption Positioning	F	<ul> <li>Update por automatically</li> </ul>			
# Panel © Plot		Enter plot name to save	<b>*</b> 5av	e plot	
Y Axis scale:	Y Axis tick format:				
e Linear	* default	Clicked points			
D Log10	O Comma separated	Brushed points			
	<ul> <li>Percett</li> </ul>	Ponta, Lines ColoriGeoup/Split/Size/Fill Mappings Boxplots Histograms/Demity/Bar			
A ANS SCARE	X Axis tick format:	Quantile Regression Smooth Linear Logistic Regressions Mean (C), Median (Pis) Maclan Merer (C)			
# Log10 () Log 10'x Format					
		Conelation Coefficient	Text Labels Rug Marks		

Cytel

### **Example tools: Meta-Analysis App**

- The Meta-analysis app is designed to provide analysts with simple steps to walk through a meta analysis
- It assumes that the user has some statistical training
- There are two simple steps
  - 1. Select studies with a common treatment and outcome of interest
  - 2. Specify a meta analysis by selecting subgroups to analyze and your meta analysis parameters
- The results from selected studies are analysed and output is provided in the form of plots and tables
  - By default, a study-level meta analysis is returned
  - Subgroup analyses can be performed by selecting subgroup variables to analyse
  - The analyst can also select which time point to look at, as well as specifying several meta analysis parameters





### **Federated Data sharing & Analysis**

- Provides the ability to improve dataset accessibility and usage
- Whilst retaining governance and control
- Reducing data movement
- Reducing risk



Your Workspace enables federated data sharing and analysis

#### In essence:

- Send requests from the trusted research environment Workspace to Data custodian(s) and have responses returned
- These responses could be
  - a) Complete or partial data sets
  - b) Groups of records
  - c) Result of analytics requests executed in the Data custodian's environment (Federated Compute)

### **Common API for Federated Data sharing**

- Working to establish an **open standard** for data platforms to participate in data sharing networks
  - Implement once, join multiple networks
  - Encourage convergence of existing (proprietary or niche) efforts
  - Encourage an ecosystem of tools & syndication
- Use existing standards where they fit
  - OpenAPI for documenting the API
  - OAuth2 for API authentication
  - W3C DCAT standard for catalogues
  - GA4GH TES = Task Execution Service for compute tasks
- Develops/adapts to specifics in clinical research
  - Data dictionary
  - Adapt TES for structured data (it's usually used for genomes, images)



OpenAPI definitions for the Common API for Federated Data Sharing are on github

U



	Partners
1	he Common API is an open source co-development between a number of partner orga
<	ADDI Alzheimer's Disease Data Initiative
	ari <mark>chia DRE</mark>
	arichia DRE Acknowledgments
1	Acknowledgments he Common API gratefully builds on work from standardisation communities:
1	Acknowledgments the Common API gratefully builds on work from standardisation communities: • World Wide Web Consortium (W3C)
1	Acknowledgments the Common API gratefully builds on work from standardisation communities: • World Wide Web Consortium (W3C) • GraphQL Foundation
1	Acknowledgments The Common API gratefully builds on work from standardisation communities:  • World Wide Web Consortium (W3C) • GraphQL Foundation • Global Alliance for Genomics and Health (GA4GH)

The code is licensed under the Mozilla Public License 2.0 see LICENSE

### Federated Data Sharing API - a Software Development Kit

- DIGITAL RESEARCH ENVIRONMENT
- Approach taken is to create an open ecosystem through a software development kit (SDK) approach
  - Data Partners can implement the API in a way that suits their local circumstances
  - Data owners retain control of permissions on their data
  - Researchers can implement their analysis plans or develop tools that suit their needs on the Workbench
- Currently, the Common API is an open specification shared by multiple initiatives (HDR UK, ADDI, ICODA)
  - <u>https://github.com/federated-data-sharing/</u>
  - Promote sharing of tools developed on this specification
- Sample code, features and tools built in the Workbench based on user feedback





### **Results & Output checking**

- Your Workspace is a trusted research environment with Airlocks
- Option for Research reviews to provide advice and support to you as a researcher, or to allow a data contributor to comment
- Results are subject to an output checking process
  - Intent is to reduce disclosure risk to ensure 'Safe Outputs'
  - May be carried out by ICODA or member of Stats Expert Group dependent on need



### **Onboarding and Training**

Our model is self-service, bring your own tools and train the trainers, but we offer a range of approaches for getting researchers productive.



ICODA onboarding session with basic Workbench training

Small Group training for research teams as required

#### Online Training Courses

Workspace Fundamentals Workspace Advanced Topics Workspace Tenant Admin Workspace Admin

FAIR Fundamentals FAIR Advanced Topics FAIR Admin Q+A drop in sessions -Open

Project team only

Aridhia Knowledge Base

Online assistance

Worked examples and demos on github

### Support



#### Contact hours

• 09:00 – 20:00 (UK) Monday-Friday

Service Level	Response Time	Target for resolution/ workaround
Severity 1	1 hour	4 hours
Severity 2	2 hours	8 hours
Severity 3	2 hours	16 hours
Severity 4	2 hours	75 hours

#### **Methods of Contact**

- Email: <a href="mailto:servicedesk@aridhia.com">servicedesk@aridhia.com</a>
- Portal: <u>https://aridhia.atlassian.net/serviced</u> <u>esk/customer/portals</u>
- Phone: +44 (0) 131 560 1480

#### Service Levels

- Workbench is operational 24x7
- 99.5% service availability (Excludes planned downtime ... platform upgrades, security patching)

### **Security Overview**



- Information security is at the heart of everything Aridhia does from the start of the Software development lifecycle right through to customer support.
- All employees are provided with security training when joining and refresher training on an annual basis.
- Aridhia operates a transparent and supportive security program that empowers it employees to report issues, suggest improvements and is continually reviewed.

### **Certifications**

Aridhia completed ISO 27001 certification in June 2019 and has maintained this certification through multiple audits. Aridhia also holds several further UK certifications and is now working to achieve HITRUST certification.



## **General Data Protection Regulation**



Aridhia achieves compliance with GDPR through the implementation of policies and processes which ensure that:

- Information is processed on a lawful and transparent basis. •
- Strong data security is achieved through design. •
- Information security governance and accountability within Aridhia is clear. •
- Individual privacy rights are respected. •

#### **Software Development Lifecycle**

In developing the DRE, Aridhia follows the OWASP Top 10 guidelines and uses tools to ensure our software complies with the OWASP best practice framework and that a "security by design" approach is followed.

Measures in place to ensure we follow a secure software development process, include:

- Coding controls are implemented. ۲
- Privacy Impact Assessments are conducted.
- Frequent regression tests, both automated and manual ensure any work for new features does not introduce security flaws.
- Separate and secured development and test environments.
- Vulnerability scanning process. ۲
- Regular penetration tests are conducted by independent security companies.



### **Workbench Security Controls**



- All user access is via HTTPS URL protected by a rooted certificate issues by DigCert SHA2 Secure Server CA, utilising sha256RSA signature algorithm with sha256 signature hashing algorithm. Will only utilise TLS 1.2 protocols or above.
- Encryption in transit All internal network traffic is protected by HTTPS or, TLS 1.2 or higher protocols.
- Encryption at rest By default, Microsoft AZURE encrypts data using FIPS 140-2 compliant 256 AES encryption for storage accounts and virtual machine disks.
- Two-factor authentication is required to access DRE services.
- The secure Workspace boundary is created through a virtual network configuration and enforced through a permissions model.
- An Intrusion Detection System and Intrusion Protection System is implemented with security alerts automatically raised to Aridhia's Service Desk Team.
- Data upload and data extraction is only permitted through an approval process.
- All uploads go through a malware scanning process.
- Full audit reporting of events.

Provided by Aridhia Informatics arithm

