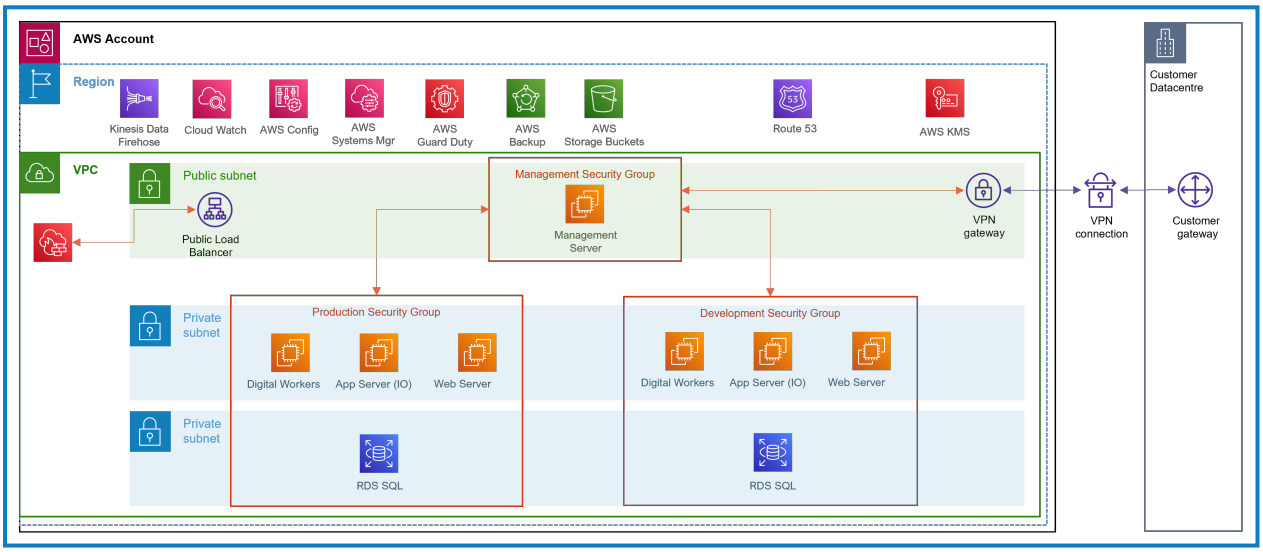
Data Security

This guidance details the transition and storage of client data through the Blue Prism® Cloud Digital Workforce platform. The guidance is intended to clarify precisely how client data is used within the platform and where, if applicable, client data is held so that an informed decision can be made regarding the protection of private and sensitive data.

For each of the applicable components in the Blue Prism Cloud Digital Workforce platform, details of data transition and storage are specified. This is performed through responding to a series of questions to explain how data is managed through the platform. These questions are typical of client and partner requests that Blue Prism Cloud have received.

The diagram below illustrates the components within the platform architecture hosted on AWS.



During the deployment of the Blue Prism Cloud platform, Blue Prism do not require access to live client data. It is recommended that only anonymized test data is used which does not show any private or sensitive information. However, should this not be possible Blue Prism have operational controls in place to ensure client data is controlled in accordance with the policies and procedures that are evaluated as part of our ISO27001:2017 accreditation and our role as a data processor according to GDPR.

This document sets out to explain how client data is used within the Blue Prism Cloud Digital Workforce platform and to understand the platform controls that are in place. Client data which is processed within the applications and systems is intended to be never held at rest within the environment, however due to the nature of some clients business processes this is not always possible. The various platform components handle information differently and have distinctive safeguards in place to minimize, if not stop, the storing of client data when / if they are used.

Security recommendations

The following security recommendations should be considered when working with the Blue Prism Cloud Digital Workforce platform:

Area	Recommendation
Blue Prism	In line with Blue Prism best practices, it is recommended that logging levels should be set to Errors Only for production environments, generating limited logs only when an issue presents itself. If required, logging levels can then be changed to help diagnose the issue. Client operators are in control of logging levels and can change them though this is not recommended except in the development environment where clients should use non-sensitive information during development.
OCR	Documents should be stored in the client Folder Store/File Share and not within the platform itself. Folders are then monitored and once populated, the OCR component will ingest and issue the captured fields to a queue in the RPA capability.
IADA.ai	When consuming IADA.ai, information is processed externally from a client platform, due to IADA.ai leveraging third-party cognitive services. Users of the Digital Workforce platform are in control as to what information is sent to IADA.ai within an automation.

Components

The sections below provide a question and answer style approach to explain the flow of data within the Blue Prism Cloud Digital Workforce. All aspects of the platform components are included, with the document order being aligned to the event timeline journey of work (business process tasks) execution.

Blue Prism

The following are a set of questions that are asked when inquiring about data flow and storage for Blue Prism.

Q1. When a digital worker migrates data from one application into another, where is the transient information held?

When a digital worker migrates data from one application to another as part of an automation, information (data) will exist in the memory of the digital worker executing the task. The information will be purged through the following methods:

- If a copy and paste is performed, then any new copy will override the previous value.
- If the data is held in memory the default Windows memory management function will control clean-up. In addition, digital workers should be scheduled for a restart every 24 hours by a client operator.

Q2. When an item is written to a Queue, what data is held?

Data items and timestamps for audit purposes are stored against the work queue item and held within the RPA Database.

Q3. When are work Queue items deleted?

Queue items lifespan are controlled by the operators of the platform, recommended best practice direct practitioners to clear these out at the end of a process or at least as part of a daily cleanup activity.

Q4. When an item is pending completion, where is the data associated with the item held?

These are held within the session logs within the RPA Database.

Q5. When an item is marked as completed, what happens to the data collected as part of the automated process?

These are held within the session logs within the RPA Database.

Q6. When an item is marked as completed, what happens to the original data received?

These are held within the session logs within the RPA Database.

Q7. When a digital worker completes a process, what information is logged?

Blue Prism Cloud strongly recommend that 'Disabled' is the logging level that is set in the Production environment to minimize the risk of sharing private data.

Q8. In Development what is the logging level?

In Development, Blue Prism Cloud recommend 'Errors Only' as the default logging level across the environment. This can be increased by the client during development if needed by setting the logging level for key Stages only. It is not recommended to set all Stages to ‘Enabled’ as prolonged periods of increased logging will cause a degradation in service.

Q9. In Production what is the logging level?

In Production, Blue Prism Cloud recommend 'Disabled' as the default logging level across the environment. Logging should not be increased as this control ensures data items or information of interest is not inadvertently collected within the RPA database.

Q10. What information is permanently stored within the Blue Prism RPA database?

The following items are permanently stored:

- Process Workflow Logic
- Application Credentials (optional / where relevant to the business process)
- Environmental Variables, for example, a path to the location of a file or a URL
- Process log information, dependent on logging level set by the client operator

Q11. How are the encrypted credentials provided to a digital worker?

The Blue Prism Application Server retrieves the encrypted credential from the database and then decrypts the returned value before issuing the credential and process to a digital worker. The Blue Prism Application Server to digital worker channel is encrypted using WCF: SOAP with Transport Encryption.

Q12. What encryption techniques are used for the Blue Prism database?

Transparent Data Encryption and AES256 bit encryption. Data in transit is encrypted with Transport Layer Security (TLS).

Q13. What options are available for the deletion of log information from the Blue Prism database?

The logging levels recommended for Production and Development are configured to protect against the inadvertent collection of information. Should the client wish to purge the logs it is possible through manipulation of the SQL database performed by Blue Prism Cloud Support via an agreed Change Request.

Authentication Server

The following are a set of questions that are asked when inquiring about data flow and storage for the Authentication Server component.

Users access the Hub and Interact environments through Authentication Server. Authentication Server contains the configuration settings for connecting to your organization's Blue Prism environment. Authentication Server therefore stores the following information in the Authentication Server database :

- Connection strings to the platform
- LDAP connection settings (if configured) including Group Role mappings
- User credentials and roles assigned
- Configuration details of service accounts
- Synched users between Authentication Server and Blue Prism (if using Blue Prism version 7.0)

Q1. What form of encryption does Authentication Server use?

Users log in through Authentication Server which determines the components they have access to. Authentication Server uses SHA-256 to hash the client secret and client ID.

Password storage – The AspNetIdentity library is used for password hashing and uses the following algorithms:

- PBKDF2 with HMAC-SHA256
- 128-bit salt
- 256-bit subkey
- 10000 iterations

Hub

The following are a set of questions that are asked when inquiring about data flow and storage for the Hub component.

Q1. What information is stored in Hub?

Hub is primarily a presentation layer of information within the platform and as such does not store a copy of information from Blue Prism. Hub does however store the following items of information:

- Dashboard details, including details of widgets and saved configurations
- License details
- Business process metadata, for example, Process Name, Priority, SLA, and so on
- Email settings
- Plugins installed
- Automation Lifecycle Management (ALM) details, Including Wireframes, Process Definitions, Exceptions and Applications (if installed and licensed)
- Interact form details (if installed and licensed)

The Hub installer generates a PFX certificate and saves it to Trusted Root Certificate Authorities. The Identity Server uses it to encrypt the JWT token and to validate the license file. The JWT token is encrypted by the RSA-SHA-256 algorithm and the key size is 2048 bit.

Q2. What log information is held in Hub?

In Hub, the following information is logged and presented on the Audit tab:

- Added user
- Edited user
- Retired user
- Made live user
- User login
- Failed login
- User logout
- SMTP settings edited
- Added role
- Edited role
- Deleted role
- Made live authentication
- Edited authentication
- Deleted authentication
- Authentication synced
- Retired authentication
- Added authentication
- Forms submitted
- Forms approved

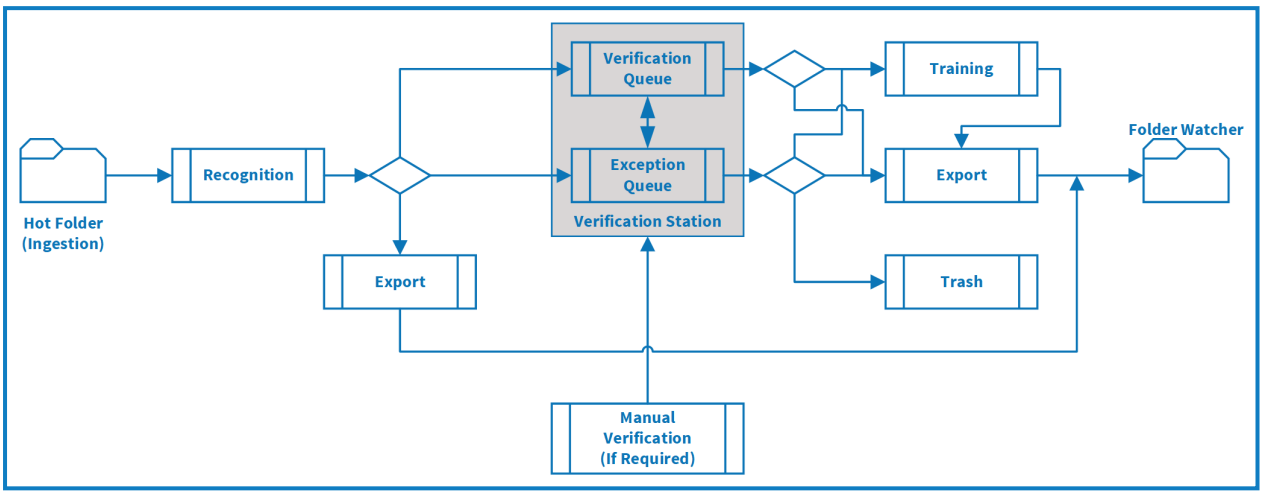
- Forms declined
- Create service account
- Edited service account
- Deleted service account
- Regenerated secret for service account
- Change password
- Created business process
- Reactivated business process
- Retired business process
- Deleted business process
- Created process definition
- Removed process definition
- Deleted process definition
- Increase major – audit logging for increasing the process definition major version
- Sign off started
- Sign off process definition
- Redacted process definition
- Owner redacted process definition
- Sign off finished
- Update to Active Directory users
- Active Directory users created
- Active Directory synchronization started
- Active Directory synchronization completed
- Active Directory synchronization failed
- Active Directory domain added

Q3. Where is the log information stored in Hub?

Log information is stored in the Audit database.

IADA OCR

The following are a set of questions that are asked when inquiring about data flow and storage for the IADA OCR component. The data flow diagram for IADA OCR is illustrated below:



Q1. Where does the input file exist?

The file exists in the client local share outside of the Digital Workforce platform. This is referred to as the 'Hot Folder'.

Q2. How does a document get into the Hot Folder?

It is moved / copied there by either an end user or a digital worker. This is a process dependent decision.

Q3. What happens to the document within the Hot Folder after processing?

Blue Prism Cloud standard configuration is to delete the document immediately after successful ingestion and it produces a temporary XML copy. The original document is managed via the Image Import Profile settings. Documents can be configured to be moved into a dedicated subfolder within the same Hot Folder location in the client local share if required.

Q4. When processing the file (from the Hot Folder), where does the temporary XML copy exist?

This is the process between the Hot Folder and Recognition activity in the diagram above, where a copy is created and stored on a dedicated disk on the Blue Prism Application Server.

Q5. When is the temporary XML copy removed from the database?

This is deleted upon export.

Q6. What data is extracted from the file?

A Document Definition template is used to identify the data from the document that is to be extracted. This data is extracted into the temporary XML.

Q7. How long is the extracted data in the Document Definition held for?

By default, all processed batches are deleted immediately after export. This can be changed by the client to 14 days if required, though it is recommended that delete immediately is used when handling sensitive data.

Q8. Where does the data extracted as per the Document Definition exist?

The temporary XML that is created and has the extracted data within it, is stored in the OCR database or on the Blue Prism Application Server disk (configurable by support based on usage load) until deleted (see question 7).

Q9. What happens to the data once it has been processed?

If the input data is processed without requiring verification, then the OCR function will output the temporary XML of the captured data items into a folder on the Blue Prism Application Server. In the diagram above this is the Recognition to Export route.

Q10. When a document cannot be automatically processed and requires verification, where does the data exist?

Information is held in the OCR database until verified.

Q11. During the Verification stage, is any data temporarily / permanently logged?

Any data that is recognized will be stored in a database until verification is complete. After which the document information is exported and the original either deleted or stored as per configuration parameters. It is recommended that the default of 'immediate deletion' is preserved.

Q12. When training is required on an item that has gone for verification, where does the data exist?

Information is stored in the OCR database and held permanently, see question 13.

Q13. What is contained within Training Data?

Training data is a full copy of a document or PDF, this includes the page structure but also the field level information. It is recommended practice that example documents are used when providing training data due to it being persistently stored.

Q14. During the Verification stage, is any data logged?

Data used during document processing is stored in the OCR database until automatic cleanup of Event log and Report Data occurs, by default it is set to:

- Event Log – deleted automatically after 14 days
- Report Data – deleted automatically after 180 days

No metadata is stored in these logs.

Q15. In the Verification stage where the item is 'Trashed', what happens to the item?

All document data is permanently removed from the OCR database. Event logs and Report Data will remain; however, no document data is kept.

Q16. During the Verification stage where an item is sent to 'Training', is any data temporarily / permanently logged?

Any document that is sent to Training will be stored as a whole document, see question 13. It is recommended that the training batch is set to "Lock Training by Operators" in the training batch to stop Operators inadvertently sending sensitive data to the OCR database.

Q17. What approach is used to read/extract the XML file outputted by the OCR function?

The XML file (housed on the Blue Prism Application Server as above) is polled by FolderWatcher, a Blue Prism Cloud proprietary Windows Service to extracts the data and send to the IADA web service.

Q18. Does FolderWatcher log any of the extracted data?

No.

Q19. Does the IADA web service log any of the extracted data?

No.

Q20. What happens to the XML file created by the OCR function, once FolderWatcher has extracted the data?

FolderWatcher will delete the temporary XML file once it has successfully passed the data to the IADA Web Service.

Q21. How does the data received by the IADA web service present itself into a Queue?

IADA creates a Queue item record which included the collected data in an Blue Prism collection.

Q22. What Management Information or Report Data is stored within the platform?

Report Data information stores the following:

- Processing time of a document
- Number of documents processes
- Stages navigated (such as, Verifications, Exceptions, Training, Exports and Trashes)

IADA.ai

The following are a set of questions that are asked when inquiring about data flow and storage for the IADA.ai component covering language translation, sentiment analysis, text analysis, language detection, and extraction of key phrases.

Q1. When IADA.ai is called, what information is sent by the digital worker?

IADA.ai is a service that takes one or many inputs and produces one or many outputs. For example, the input may be an email body with the second input being a target language the message should be translated into. In this scenario, both the email body and the specified language choice would be sent by the digital worker. This information is not obfuscated and therefore the sending of personal data and information has to be controlled by the organization.

Q2. What information is logged by the digital worker when calling IADA.ai?

The logging levels set by the RPA would be used when the digital worker calls IADA.ai, please refer to the Blue Prism section of this document.

Q3. Where does IADA.ai reside?

IADA.ai resides within a centralized Blue Prism Cloud Microsoft Azure subscription which is external to any client or partner platform subscription or account.

Q4. Where is the information processed?

Information is processed within the IADA.ai web application which resides within the Blue Prism Cloud Microsoft Azure subscription – thereafter, a Microsoft Cognitive service is called which also processes the sent data by IADA.ai. The Microsoft Cognitive service does not log nor store any information other than a record of the event.

Interact

The following are a set of questions that are asked when inquiring about data flow and storage for the Interact component.

Q1. How is data added to an Interact form?

Forms are populated by a user or digital worker with the appropriate levels of access with text data and / or an attached file.

Q2. When a form has been completed but not submitted (so is in a draft state) where does the data exist?

The data is held within Interact, be it a text value, numerical value or file attachment, this is stored within the Interact database. If the **Purge data on submission** option is set for a form, data can only be purged once the form has been submitted.

Q3. What log information is held in Interact?

See question 2 in the Hub section for more details on the audited information.

Q4. Where is the log information stored in Interact?

Log information is stored within the Audit database.

Q5. When a form has been completed and submitted where does the data exist?

The data is stored in the Interact database for auditability purposes. The data may also be stored in the RPA database and any end-point databases defined within the automation.

Fields/attachments marked as sensitive (using the **Purge data on submission** functionality) are deleted from Interact.

Q6. How is a submitted form presented to a Queue for execution by a digital worker?

Interact 'calls' the IADA web service, sending all collected data.

Q7. Does the IADA web service log any of the extracted data?

No.

Q8. When a digital worker sends information to Interact, where is the data stored?

All data sent by the digital worker will be stored against the Interact submission (within the Interact database).

Q9. Does Interact utilize any encryption to protect its data?

Interact supports encryption as detailed below:

- **Traffic encryption** – Enable HTTPS only communication for production. Requires customers to provide TLS certificates for all web applications and all communication channels must be secured.
- **Data protection** – The Hub installer generates a PFX certificate and saves it to Trusted Root Certificate Authorities. All applications use it to encrypt sensitive data, such as connection strings in the appsettings.json file. Data protection uses the following default algorithms:
    - Encryption Algorithm is AES-256-CBC
    - Validation Algorithm is HMACSHA256
    - The key size is 2048 bit.

RabbitMQ

RabbitMQ is an open-source message broker which supports multiple messaging protocols. RabbitMQ is used within the Blue Prism Cloud platform architecture to manage message queuing, passing the messages from one system to another.

Messages only remain within RabbitMQ until the receiving system can handle the request. Once the request is received, the message is removed and nothing is stored in the RabbitMQ queues.

By default, messages are encrypted using a standard TLS connection and starting from the 4.4 version of the platform an additional layer if client encryption has been included utilizing AMQPS (TLS/SSL encrypted AMQP).

For more details about RabbitMQ and how it is used, please see the RabbitMQ use in Hub and Interact help page.