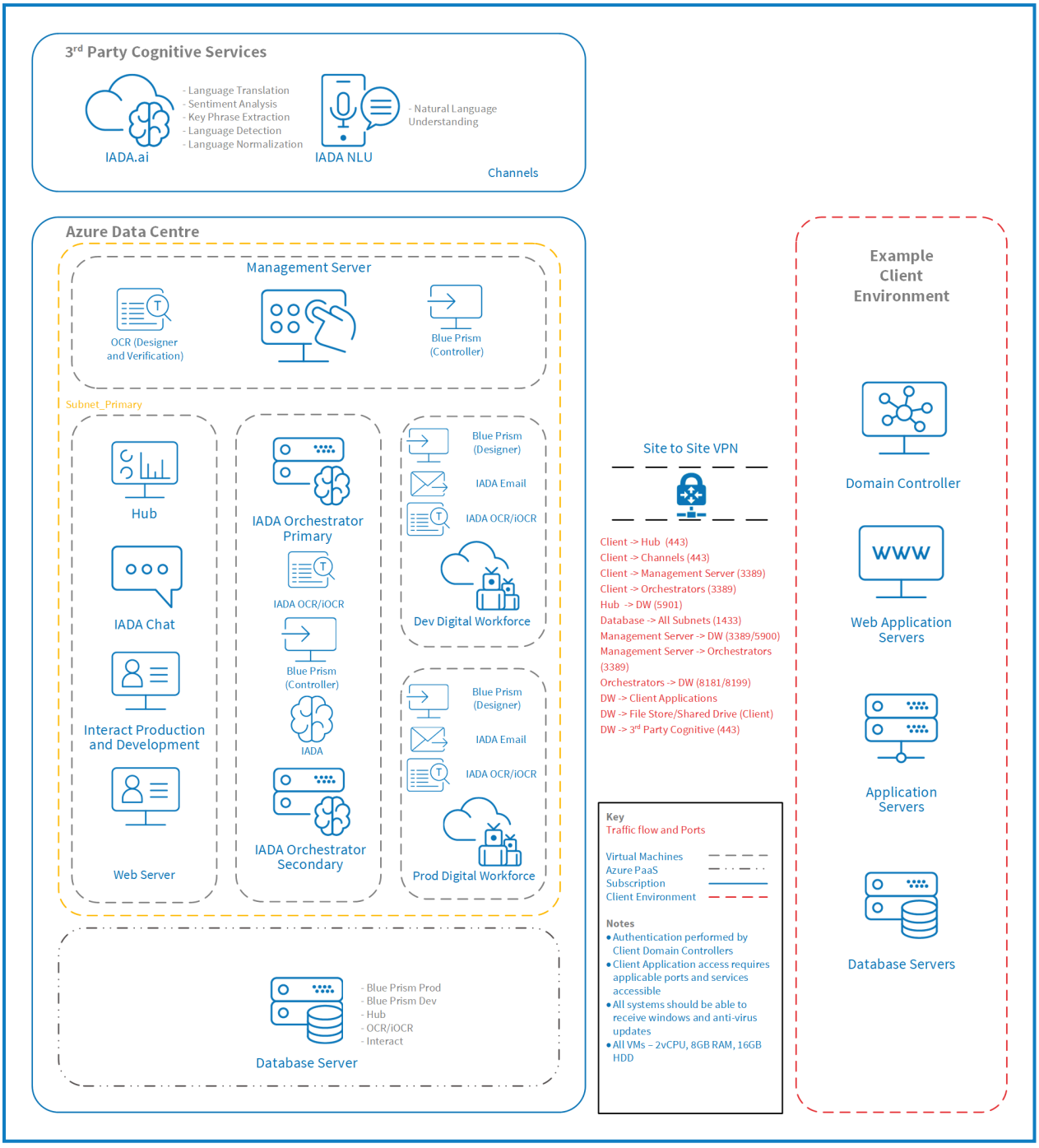
Data security

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 a 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 typical architecture and can be used to reference the components within the platform architecture.



During the deployment of the Blue Prism Cloud platform, Blue Prism Cloud do not require access to 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 Cloud 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.

For example, Interact contains a 'Purge on Submission' feature for fields on all forms, when selected at the point the form is submitted all entered values in those fields are purged from the Interact instance.

Should a form be saved as a 'Draft' then any field with 'Purge on Submission' enabled, the data that has been entered into the field will be removed. However, if the form is submitted and the form is set for 'Approval', the populated form will hold a record of the information until the Approver submits the form.

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 disabled for production environments, negating any information being collected and stored at rest. 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	When consuming IADA.ai, information is processed externally from a client platform, due to IADA.ai leveraging 3rd 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

This sections below provide a question and answer style approach to explain the flow of data within the Blue Prism Cloud Digital Workforce, Evolution Edition Release 4. 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 enquiring about data flow and storage for Blue Prism.

- **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.

- **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.

- **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.

- **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.

- **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.

- **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.

- **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.

- **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.

- **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.
- **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, e.g. a path to the location of a file or a URL;
  - Process log information, dependent on logging level set by the client operator.
- **How are the encrypted credentials provided to a digital worker?**

  The IADA Orchestrator retrieves the encrypted credential from the database and then decrypts the returned value before issuing the credential and process to a digital worker. The IADA Orchestrator to digital worker channel is encrypted using WCF: SOAP with Transport Encryption.
- **What encryption techniques are used for the Blue Prism database?**

  Transparent Data Encryption and AES256 bit encryption.
- **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.

Hub

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

What information is stored within 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 for example. Hub does however store the following items of information.

- Dashboard
- Wireframes:
  - Digital worker meta data
- Hub User Credentials
- Connection Strings to the platform
- License Details:
  - Business Process meta data, e.g. Process Name, Priority, SLA, etc.
  - Email Settings

IADA Chat

- **When communicating with IADA Chat, where is my conversation being processed?**

  The conversation is processed in a 3rd party application for example, Skype, with communications being sent back to the digital worker which utilizes IADA NLU for classification. IADA NLU leverages Azure Cognitive Services where Microsoft's NLP is accessed to deconstruct the meaning of the naturally formed text.
- **What information is sent to IADA Chat when conversing?**

  All text is being sent to IADA Chat during an interaction with the conversational user interface.
- **Where does IADA Chat reside?**

  IADA Chat resides as an Azure Function App within a client or partner subscription.
- **For the Natural Language Processing aspect of IADA Chat, where does this processing take place?**

  IADA Chat leverages IADA NLU and Microsoft Azure Cognitive Services. All processing takes place external to the client or partner subscription within the Microsoft Azure Cognitive Services.
- **What information is logged by IADA Chat?**

  Event and session information is stored by IADA Chat however the conversational inputs are not stored nor logged.
- **What training data is required for IADA Chat to understand a conversation?**

  Utterances which are made up of historic scenarios make up the training data. Training data should exclude sensitive information as this has no bearing on the confidence.
- **What information is stored in IADA Chat?**

  IADA Chat does not store a record of conversations or message details, however, does record the transaction for audit purposes. Note: The client application used to communicate with IADA Chat, e.g. Skype, may store a record of the conversation history on the client side, this is configurable on a client application basis.
- **What information is stored in IADA NLU?**

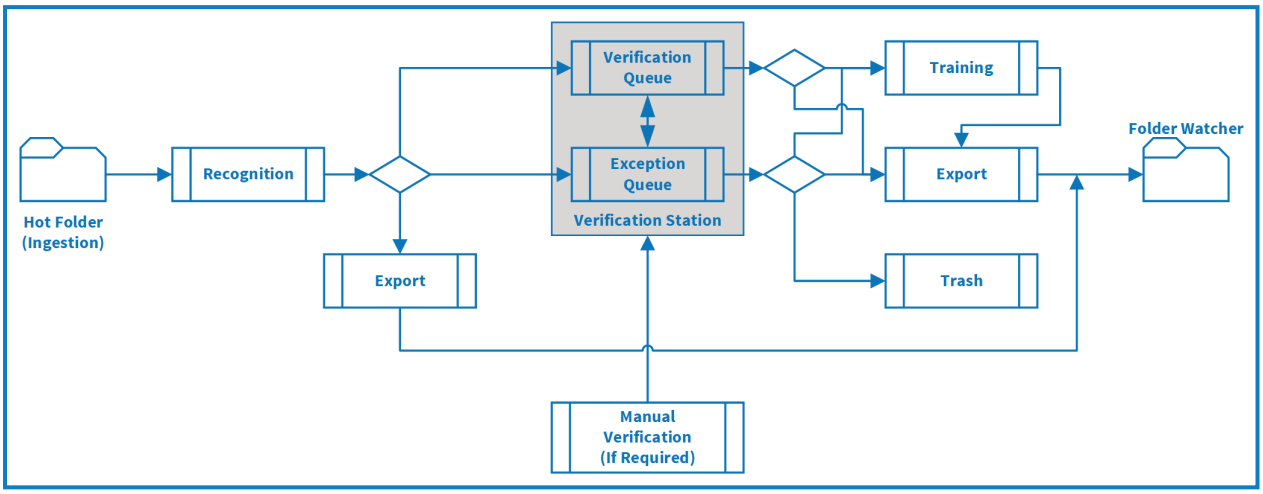
  IADA NLU requires the following information to understand and recommend intents when being invoked:
  - Intent names
  - Utterances
  - Entities
  - Records logging information for audit purposes.
- Messages sent to IADA NLU are not logged or stored at rest.
- **What information is stored in Azure Cognitive Services?**

  A record of the transaction is stored for audit purposes, however no client data nor submitted conversational information is stored.
- **Where is the training data stored which IADA Chat uses to understand a conversation?**

  Training data is stored across the Blue Prism Cloud IADA NLU service as well as within a Microsoft Cognitive Service repository – external to the client or partner subscription.

IADA OCR

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



- **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'.

- **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.

- **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.

- **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 IADA Orchestrator server.

- **When is the temporary XML copy removed from the database?**

  This is deleted upon export.

- **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.

- **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.

- **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 IADA Orchestrator disk (configurable by support based on usage load) until deleted (see question 7).

- **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 IADA Orchestrator server.In the diagram above this is the Recognition to Export route.
- **When a document cannot be automatically processed and requires verification, where does the data exist?**

  Information is held in the OCR database until verified.
- **During the Verification stage, is any data temporarily / permanentlylogged?**

  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.
- **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.
- **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.
- **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.

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

  The XML file (housed on the IADA Orchestrator 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.
- **Does FolderWatcher log any of the extracted data?**

  No.
- **Does the IADA web service log any of the extracted data?**

  No.

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

  IADA created a Queue item record which included the collected data in an Blue Prism collection.
- **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 (i.e. Verifications, Exceptions, Training, Exports and Trashes)

IADA.ai

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

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

  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. Blue Prism Cloud provide an obfuscation utility which means character formats of the following formats are obscured:

  - Censoring Email Address
  - Censoring URL
  - Censoring Postcodes
  - Censoring Telephone Numbers
  - Censoring IPV4 Addresses
  - Censoring IPV6 Addresses
  - Censoring National Insurance Number
  - Censoring Social Security Number
  - Censoring All Numeric Characters
  - Censoring All Alpha Characters
  - Censoring Upper Case Characters
  - Censoring Lower Case Characters
  - Censoring All Symbol Characters
- **What information is logged by the digital worker when calling IADA.ai?**

  The loggings 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.
- **What information does IADA.ai store / log when processing a request?**

  IADA.ai stores a record for audit purposes of the event, however it does not store any of the data used in the information it received.
- **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.
- **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 enquiring about data flow and storage for the Interact component.

- **How is data added to a 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.

- **When a form has been completed but not submitted (i.e. 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. Data can only be purged once a form has been submittedif the Purge Upon Submission flag is set.

- **What Log information is held within Interact?**

  Within Interact, the following information is logged and presented within the Audit tab:

  - Successful logins
  - Logouts
  - Failed logins
  - Creation of user
  - Editing of user
  - Creation of process
  - Editing of process
  - Creation of process category
  - Submission of process
  - Updating of license key
  - Updating of SMTP settings
  - Updating of LDAP settings
  - Updating of SQL settings
  - Endpoint synchronization
  - External API calls

- **Where is the Log information stored within Interact?**

  Log information is stored within an encrypted column within the Interact database.

- **When a form has been completed and submitted where does the data exist?**

  Fields/Attachments marked as sensitive (using the Purge Upon Submission functionality) is deleted from Interact. All other fields not marked as sensitive will remain within the Interact database for auditability purposes.

- **Is there an archive or deletion feature within Interact for Request management?**

  Within Interact, a user can archive and un-archive their Historic Requests. There is however no mechanism for permanently deleting Requests submitted by users. Should the client wish to delete information it is possible through manipulation of the SQL database performed by Blue Prism Cloud Support via an agreed Change request.

- **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.

- **Does the IADA web service log any of the extracted data?**

  No.

- **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 request (within the Interact database), in the fields nominated as part of the process. The feature, 'Purge Upon Submission', cannot be invoked once a form has been submitted i.e. sensitive fields are not purged when a digital worker sends information to Interact. If sensitive data / client data is sent back to Interact from Blue Prism this will be stored in the Interact Database and as per question 6 cannot be deleted.