

Date of Approval: 05/16/2025  
Questionnaire Number: 1892

## Basic Information/Executive Summary

What is the name of your project (system, database, pilot, product, survey, social media site, etc.)?

Gaming Industry Tip Compliance Agreement Analysis & Compliance for Tip Reporting Suite

Acronym:

GITCA ACTR Suite

Business Unit

Research Applied Analytics and Statistics

Preparer

# For Official Use Only

Subject Matter Expert

# For Official Use Only

Program Manager

# For Official Use Only

Designated Executive Representative

# For Official Use Only

Executive Sponsor

# For Official Use Only

Executive Summary: Provide a clear and concise description of your project and how it will allow the IRS to achieve its mission.

The ACTR (Analysis & Compliance for Tip Reporting) Suite enables NTRCP (National Tip Reporting Compliance Program) analysts to process GITCA (Gaming Industry Tip Compliance Agreement) forms, perform tip rate calculations, store extracted data, and generate/view analytics. The ACTR suite is made up of a web application developed using Plotly Dash (a Python framework) for data ingest and modification, a PostgreSQL database for data storage, and a series of Power BI Reports for data consumption, reporting and analytics. This project will allow the IRS to achieve its mission by significantly decreasing the time it takes to ingest, process, and perform calculations on GITCA forms. This will allow NTRCP analysts to reduce their backlog and initiate more GITCA agreements, which will reduce the tax gap.

## Personally Identifiable Information (PII)

Will this project use, collect, receive, display, store, maintain, or disseminate any type of Sensitive but Unclassified (SBU), Personally Identifiable Information (PII), or Federal Tax Information (FTI)?

Yes

Please explain in detail how this project uses sensitive data from inception to destruction (data lifecycle).

Users (NTRCP analysts) will upload tax forms and taxpayer data into the ACTR Suite, it will read these files (sometimes with OCR (Optical Character Recognition)), it will perform light processing on the read data. Then Tip rate calculations will be executed on the information extracted and the result of those calculations, as well as the information pulled from uploaded documents, will be uploaded to a database dedicated to the ACTR Suite. Certain users will be given access to PowerBI dashboards to view aggregations and summaries of information uploaded to the database.

Please select all types of Sensitive but Unclassified data (SBU)/Personally Identifiable Information (PII)/Federal Tax Information (FTI) that this project uses.

Address

Employer Identification Number

Employment Information

Federal Tax Information (FTI)

Geographical Indicators

Name

Non-Tax Proprietary data

Social Security Number (including masked or last four digits)

Cite the authority for collecting SBU/PII/FTI (including SSN if relevant).

PII for federal tax administration - generally IRC Sections 6001 6011 or 6012

SSN for tax returns and return information - IRC section 6109

## Product Information (Questions)

1 Is this PCLIA a result of the Inflation Reduction Act (IRA)?

No

2 What type of project is this (system, project, application, database, pilot/proof of concept/prototype, power platform/visualization tool)?

System

3 What Tier designation has been applied to your system?

2

4 Is this a new system?

Yes

5 Is this system considered a child system/application to another (parent) system?

No

6 Indicate what OneSDLC State is the system in (Allocation, Readiness, Execution) or indicate if you go through Information Technology's (IT) Technical Insertion Process and what stage you have progressed to.

Not Applicable; The OneSDLC process does not apply to the intelligent automation. OneSDLC [for our lab] is RPA bot specific and this intelligent automation does not fall into that category.

7 Is this a change resulting from the OneSDLC process?

No

8 Please provide the full name and acronym of the governance board or Executive Steering Committee (ESC) this system reports to.

Data and Analytics Advisory Group (DAAG)

9 If the system is on the As-Built-Architecture (ABA), what is the ABA ID number of the system? If this PCLIA covers multiple applications shown on the ABA, please indicate the ABA ID number(s) for each application covered separated by a comma. If the system is not in the ABA, then contact the ABA (<https://ea.web.irs.gov/aba/index.html>) for assistance.

Not Applicable

10 If this system discloses any PII to any third party outside the IRS, does the system have a process in place to account for such disclosures in compliance with IRC 6103(p)(3)(A) or Subsection c of the Privacy Act?

Yes

11 Does your project/system involve any use of artificial intelligence (AI), including virtual assistant, chat bot, and robotic process automation, as defined in Executive Order 13960 and 14110?

Yes

11.1 Describe the business process and purpose of your Artificial Intelligence (AI) and identify what system(s) or business process(es) this AI supports.

We use Optical Character Recognition (OCR) tools to extract text from and enable automated data extraction of scanned Tip Agreement PDFs. Since these scanned PDFs are essentially data stored as images, with the underlying data

represented by pixels, predictive OCR models trained to predict text values from these pixels is necessary to extract useful data from these forms.

11.2 What is the algorithm or learning method used and what database is training your AI?

To leverage OCR, we are using the Python library PyTesseract, which serves as a wrapper for Google's Tesseract-OCR Engine. Under the hood, PyTesseract provides several pre-trained Long Short-Term Memory (LSTM) neural networks that can be used to predict text from the pixels of an image/scanned PDF.

PyTesseract can also be leveraged to train text recognition LSTM models for specific use cases. For our purposes, we relied on the pre-trained models provided out-of-the-box.

11.3 How is this AI system tested and validated to ensure that the decisions or outputs are reliable (relevant to the input) and performs without biases and drifting? Note: Outputs include, User information, transfer to assistor, reports on dashboard depicting activities.

Tesseract is an OCR engine that has been in development since 1984, with the most recent updates being released November 2021. The pre-trained models have thus been internally evaluated and updated by Google developers. In terms of application to our data extraction from Tip Agreement PDFs, we validated the accuracy, completeness, and consistency of outputs for a sample of 30% of all agreements. Upon identifying errors, we manually corrected the specific erroneous elements for the sample and then applied analysis of these errors to manually comb and correct the rest of the agreement population. While errors will occur, it will be the responsibility of the user to review the automatically extracted tip agreement data and ensure it matches that which is presented in the original form.

12 Does this system use cloud computing?

No

13 Does this system/application interact with the public?

No

14 Describe the business process allowing an individual to access or correct their information. (Due Process)

This is generally not applicable to the application. In the event a correction, redress or access is required it would follow the general process in place as applicable. The data collected is from data that is submitted by various sources, employers, employees, third parties, and/or are part of a rev rule or rev proc or announcement. The information within the ACTR suite comes from various IRS Systems and forms. Those systems and forms provide the Privacy Act Notice to individuals. ACTR does not directly provide individuals the opportunity to decline from providing information and/or from consenting to particular uses of the information. Notice, consent, and due process are provided in the tax forms instructions, and pursuant to 5 USC.

15 Is this system owned and/or operated by a contractor?

Yes- IRS Owned, contractor operated

15.1 If a contractor owns or operates the system, does the contractor use subcontractors; or do you require multiple contractors to operate, test, and/or maintain this system?

Yes

15.2 What PII/SBU data does the subcontractor(s) have access to?

Address, Employer Identification Number, Employment Information, Federal Tax Information (FTI), Geographical Indicators, Name, Non-Tax Proprietary data, Social Security Number (including masked or last four digits)

16 Identify what role(s) the IRS and/or the contractor(s) performs; indicate what access level (to this system's PII data) each role is entitled to. (Include details about completion status and level of access of the contractor's background investigation was approved for.)

IRS Employees Roles: Users (Read, Write) "Headquarters analyst role" or "Fieldperson role", Managers (Administrator), Executive overseeing the GITCA program additionally has access to the "Executive Summary" dashboard in Power BI. All roles listed will have access to the web application. Managers, the executive overseeing the program, and Headquarters analyst role will have access to a more general summary dashboard in PowerBI.

Contractor Roles: Developers (Administrator) with completed background investigations.

17 The Privacy Act of 1974 (5 USC § 552a(e)(3)) requires each agency that maintains a system of records, to inform each individual requested to supply information about himself or herself. Please provide the Privacy Act Statement presented by your system or indicate a Privacy Act Statement is not used and individuals are not given the opportunity to consent to the collection of their PII.

Not Applicable. This is not a "system of record". This is merely an internal database. Individuals do not supply information about themselves.

18 How many records in the system are attributable to IRS Employees? Enter "Under 50,000", "50,000 to 100,000", "More than 100,000" or "Not Applicable".

Not Applicable

19 How many records in the system are attributable to contractors? Enter "Under 5,000", "5,000 to 10,000", "More than 10,000" or "Not Applicable".

Not Applicable.

20 How many records in the system are attributable to members of the public? Enter "Under 100,000", "100,000 to 1,000,000", "More than 1,000,000" or "Not applicable".

100,000 to 1,000,000

21 Identify any "other" records categories not attributable to the categories listed above; identify the category and the number of corresponding records, to the nearest 10,000; if no other categories exist, enter "Not Applicable".

Not Applicable.

22 How is access to SBU/PII determined and by whom?

Access for each of the roles will be distributed via BEARS (Business Entitlement Access Request System) Entitlements. Users are given the minimum set of privileges required to perform their regular and recurring work assignments; they are restricted from changing the boundaries of their access without management approval.

23 Is there a data dictionary on file for this system? Note: Selecting "Yes" indicates an upload to the Attachment Section is required.

No

24 Explain any privacy and civil liberties risks related to privacy controls.

By restricting access to users who can obtain a BEARS entitlement we can ensure access to the information ingested is only available to those who need it. Logging of all significant actions taken within the ACTR suite introduces accountability for users who do gain access. There is no way for users to change their own roles / permissions. The team is undergoing the Enterprise Security Audit Trails (ESAT) process in order to ensure necessary logging and security measures are in place.

25 Please upload all privacy risk finding documents identified for the system (Audit trail, RAFT, POA&M, Breach Plan, etc.); click "yes" to confirm upload(s) are complete.

No

26 Describe this system's audit trail in detail. Provide supporting documents.

Application logs contain UserID (credentials to access ACTR Suite), Timestamp, what pages are accessed, file upload metadata, and any query sent to the database. The logs are sent to our Postgress database and to the GITCA01 server. Additionally, database log files are sent from our Postgress Database to the GITCA01 server.

27 Does this system use or plan to use SBU data in a non-production environment?

Yes

27.1 Please upload the Approved Email and one of the following SBU Data Use Forms, Questionnaire (F14664) or Request(F14665) or the approved Recertification (F14659). Select Yes to indicate that you will upload the Approval email and one of the SBU Data Use forms.

Yes

# Interfaces

## **Interface Type**

IRS Systems, file, or database

Agency Name

Appendix A

Incoming/Outgoing

Incoming (Receiving)

Transfer Method

Secure email/Zixmail

## **Interface Type**

Forms

Agency Name

Point Of Sale (POS) Reports

Incoming/Outgoing

Incoming (Receiving)

Transfer Method

Secure email/Zixmail

## **Interface Type**

Forms

Agency Name

Tip Agreement

Incoming/Outgoing

Both

Transfer Method

Secure email/Zixmail

## **Interface Type**

Other Organization

Agency Name

Gaming Establishments

Incoming/Outgoing

Incoming (Receiving)

Secure Email MOU, Letter 5873

Transfer Method

Secure email/Zixmail

## **Interface Type**

IRS Systems, file, or database

Agency Name

Analysis & Compliance for Tip Reporting (ACTR) Development

Power BI Workspace

Incoming/Outgoing  
Outgoing (Sending)  
Transfer Method  
Secure Data Transfer (SDT)

**Interface Type**

Forms  
Agency Name  
Payroll Reports  
Incoming/Outgoing  
Incoming (Receiving)  
Transfer Method  
Secure email/Zixmail

**Interface Type**

Forms  
Agency Name  
Form 14439  
Incoming/Outgoing  
Incoming (Receiving)  
Transfer Method  
Secure email/Zixmail

**Interface Type**

Forms  
Agency Name  
Form 8027  
Incoming/Outgoing  
Incoming (Receiving)  
Transfer Method  
Secure email/Zixmail

## **Systems of Records Notices (SORNs)**

**SORN Number & Name**

IRS 34.037 - Audit Trail and Security Records  
Describe the IRS use and relevance of this SORN.  
Logging is implemented in the system and being refined with the  
ESAT team to ensure appropriate measures are put in place to  
identify the who, what, where, and when data is being accessed.

**SORN Number & Name**

IRS 42.021 - Compliance Programs and Projects Files

Describe the IRS use and relevance of this SORN.

SBU/PII is used to calculate tip rates for GITCA program participants. This data is already in the hands of NTRCP analysts / fieldpersons who are submitting it to the ACTR suite. Access to information ingested into the ACTR suite is restricted to those who will use it as a part of their regular function at the IRS and can obtain a BEARS entitlement. This information is critical to managing the GITCA program and its active and potential agreements.

## Records Retention

What is the Record Schedule System?

General Record Schedule (GRS)

What is the retention series title?

General Records Schedule (GRS) 3.1 for General Technology Management Records

What is the GRS/RCS Item Number?

Item 020

What type of Records is this for?

Electronic

What is the disposition schedule?

Audit and user logs are scheduled under General Records Schedule (GRS) 3.1 for General Technology Management Records, Item 020. IRS System Technology audit logs are maintained per IRM 5.1.25.6 in the Security Audit and Analysis System (SAAS). Audit Logs will be erased or purged from the SAAS at the conclusion of their retention period(s) as required under IRM 1.15.6. The method used for sanitization will follow NIST SP 800-88 guidelines. These will not be stored in the automation, but a SQL database.

## Data Locations

What type of site is this?

Data Gateway

What is the name of the Data Gateway?

Research Applied Analytics & Statistics (RAAS) - Strategy & Business Solutions - ACTR-Sandbox Test

What is the sensitivity of the Data Gateway?

Federal Tax Information (FTI)

Please provide a brief description of the Data Gateway.

Postgresql PBI gateway connection

What are the incoming connections to this Data Gateway?

GITCA01 server and actr\_dev database

What are the outgoing connections from this Data Gateway?

Analysis & Compliance for Tip Reporting (ACTR) Development workspace

What type of site is this?

Workspace (PowerBI)

What is the name of the Workspace (Power BI)?

Analysis & Compliance for Tip Reporting (ACTR) Development

What is the sensitivity of the Workspace (Power BI)?

Federal Tax Information (FTI)

What is the URL of the item, if applicable?

<https://app.powerbigov.us/groups/085b6376-95c6-4a65-8f27-db8c4eba0349/list>

Please provide a brief description of the Workspace (Power BI).

PowerBI Dashboard where approved users can access approved data visualizations on GITCA data ingested into our database.

What are the incoming connections to this Workspace (Power BI)?

Postgres Database - Actr\_dev

What type of site is this?

System

What is the name of the System?

Actr\_dev Database

What is the sensitivity of the System?

Personally Identifiable Information (PII) including Linkable Data

Please provide a brief description of the System.

PostgreSQL Database holding data ingested into the ACTR Suite and Tip rate calculations, hosted on the GITCA01 server.

What are the incoming connections to this System?

Database receives data extracted from documents uploaded to the ACTR Suite and tip rate calculations made based on that data.

What are the outgoing connections from this System?

Postgresql Power BI Gateway Connection (connects database to PBI dashboard)

What type of site is this?

System

What is the name of the System?

GITCA02 Server

What is the sensitivity of the System?

Federal Tax Information (FTI)

Please provide a brief description of the System.

Internal IRS production server that hosts the database and web application - is used only for the GITCA program

What are the incoming connections to this System?

DBeaver, SecureCRT

What are the outgoing connections from this System?

Power BI Workspace

What type of site is this?

System

What is the name of the System?

GITCA01 Server

What is the sensitivity of the System?

Federal Tax Information (FTI)

What is the URL of the item, if applicable?

10.207.92.8

Please provide a brief description of the System.

Internal IRS development server that hosts the database and web application - is used only for the GITCA program

What are the incoming connections to this System?

DBeaver, SecureCRT

What are the outgoing connections from this System?

Power BI Workspace

What type of site is this?

System

What is the name of the System?

Actr\_prod Database

What is the sensitivity of the System?

Federal Tax Information (FTI)

Please provide a brief description of the System.

PostgreSQL Database holding data ingested into the ACTR Suite and Tip rate calculations, hosted on the GITCA02 server.

What are the incoming connections to this System?

Database receives data extracted from documents uploaded to the ACTR Suite and tip rate calculations made based on that data.

What are the outgoing connections from this System?

Postgresql Power BI Gateway Connection (connects database to PBI dashboard).

What type of site is this?

System

What is the name of the System?

SPLUNK

What is the sensitivity of the System?

Personally Identifiable Information (PII) including Linkable Data

What is the URL of the item, if applicable?

<https://laco-irs.mgt.afsiep.net/>

Please provide a brief description of the System.

SPLUNK is a software platform used for searching, monitoring, and analyzing data in real-time. It helps organizations gain insights from logs, events, and other data sources by providing powerful tools for data indexing, visualizations, and reporting.

What are the incoming connections to this System?

UiPath sends log files through an integration with IEP SPLUNK.

What are the outgoing connections from this System?

IEP SPLUNK sends data through an outgoing connection to IRS SPLUNK.