

INFORME TÉCNICO PREVIO DE EVALUACIÓN DE SOFTWARE

N° 0001-2014/5E2400

1. **NOMBRE DEL AREA:**
DIVISION DE DESARROLLO DE SISTEMAS ANALITICOS
2. **RESPONSABLE DE LA EVALUACIÓN:**
RONWELD BAYONA MACHADO
3. **CARGO**
PROFESIONAL DE LA DIVISION DE DESARROLLO DE SISTEMAS ANALITICOS
4. **FECHA**
10 de julio de 2014
5. **JUSTIFICACIÓN**

La SUNAT para el logro de los objetivos estratégicos institucionales requiere de la realización de acciones que resulten de una adecuada y oportuna toma de decisiones en base a información. En la medida que los datos tengan las características de ser fiables, válidos, disponibles, oportunos, consistentes, verificables y seguros; SUNAT brindará mejores servicios a los contribuyentes y será más efectivo en el control del cumplimiento.

Las áreas usuarias para lograr las características de estos datos, realizan esfuerzos con diferente periodicidad (diario, semanal, mensual y anual), mientras esto se mantenga siempre se tendrá problemas con la calidad de los datos, los usuarios han ideado soluciones temporales y parciales para poder llevar a cabo sus funciones y procedimientos, haciendo uso de herramientas como hojas de cálculos o manejadores de base de datos temporales para tal propósito, demandando mucho tiempo y recursos técnicos, ayudando a que el problema se siga dimensionando; muchas de estas herramientas debido a su naturaleza de no manejar grandes volúmenes de información requieren mucho tiempo en horas e incluso días para procesar la información. Todo este trabajo de limpieza de datos no obedece a un procedimiento definido por lo que la data origen nunca está limpia para otras áreas, solo se limita al área y personas que realiza tal acción, duplicando esfuerzo por otras áreas que muchas veces necesitan hacer uso de la misma.

Como no existe un proceso definido no se manejan estándares ni políticas que rijan el desarrollo y la utilización de los datos a nivel institucional, tampoco hay responsables ni una estructura organizacional ad-hoc que se preocupe de la integridad de los datos y los esfuerzos se materializan en gran cantidad de tiempo como podemos ver lo siguiente: el proceso de cobranza y control de deuda demanda el 80% del tiempo del personal dedicado a limpieza de datos; el proceso de fiscalización requiere el 20% del tiempo del personal; la distracción del tiempo de

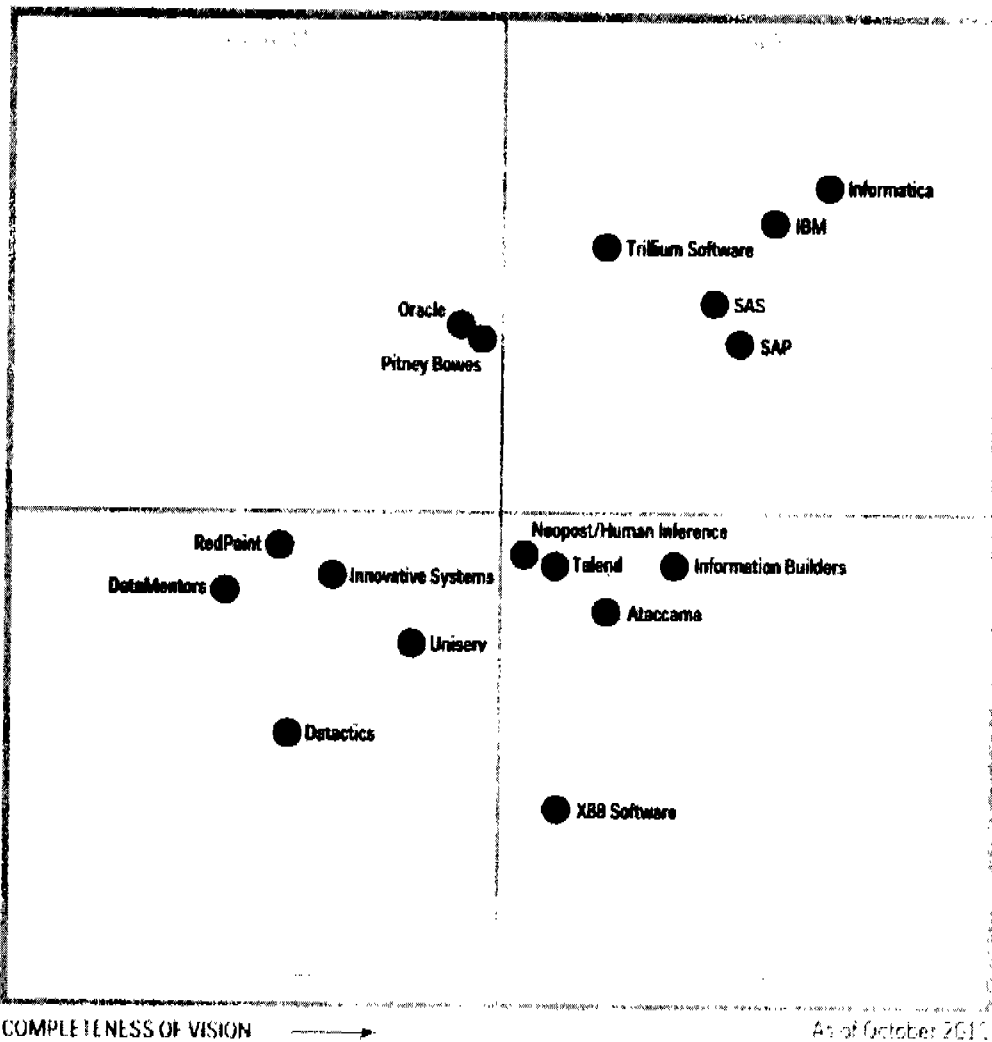


estas personas se debería aprovechar mejor en nuevas oportunidades de sus funciones.

En este contexto, se plantea adquirir una herramienta de software cuya naturaleza permita hacer limpieza y calidad de datos sobre grandes volúmenes de información, que realice perfilamiento de los mismos, cree reglas del negocio, emita resultados de los procesos de perfilamiento y ejecución de reglas, explore los resultados, realice estandarización y limpieza de datos, realice auditoría a los mismos, cree indicadores, cuadros de mando, para tener una data limpia en toda la institución contribuyendo con ello ser efectivos en los servicios que se brindan.

6. ALTERNATIVAS

Para la evaluación de las herramientas de calidad de datos se hizo la revisión del Cuadrante Mágico de Gartner como se puede ver en la siguiente figura.



Source: Gartner (October 2013)

As of October 2013



Partiendo del Cuadrante Mágico de Gartner, para el análisis se han seleccionado las herramientas de Calidad de Datos ubicadas en el cuadrante de líderes. Los productos a evaluados son los siguientes:

- Trillium Software System
- InfoSphere Data Quality
- Informática Data Quality

Trillium Software System es una plataforma de Calidad de Datos que es desarrollada por Trillium Software.

InfoSphere Data Quality es una plataforma de Calidad de Datos que es desarrollada por IBM, su versión es 9.1. y forma parte de la suite InfoSphere Information Server.

Informática Data Quality es una plataforma de Calidad de Datos cuyo fabricante es Informática.

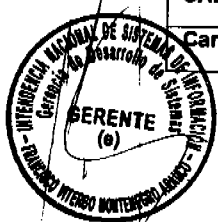
7. ANÁLISIS COMPARATIVO TÉCNICO

Para el análisis se utilizará la metodología establecida en la guía técnica sobre evaluación de software para la administración pública, aprobada mediante la resolución ministerial N° 139-2004-PCM, en el cual se establece el modelo de calidad basada en los criterios de evaluación mostrados en el anexo N° 1.

Las métricas se seleccionaron en base a las necesidades de la Institución y al análisis de la información técnica de las alternativas que se indican en el numeral 6 del presente informe.

La calidad del software ha sido evaluada, de acuerdo a un puntaje o escala fijada para cada característica, con un máximo de 100 puntos para medir la calidad total del producto. Aquel producto que no alcance el puntaje mínimo de cualquier atributo, será descartado. El siguiente cuadro muestra los resultados de la evaluación:

			Puntaje máximo	Puntaje mínimo	Trillium Software System	InfoSphere Data Quality	Informática Data Quality
PUNTAJE TOTAL			100	0	95.3	96.3	95.7
CALIDAD EXTERNA			60	0	56.3	56.5	56.6
Características	Subcaracterísticas	Requisito del software					



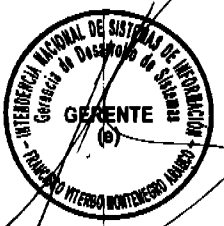
Funcionalidad	Adecuación						
		<ul style="list-style-type: none"> ➤ Cuenta con una interfaz gráfica basada en tecnología drill & drop. ➤ Permite crear proyectos o ambientes de trabajo ➤ Capacidad para organizar por proyecto o ambiente de trabajo en objetos o estructuras para un mejor control del mismos (fuentes de datos, reglas, Web Services, reportes, aplicaciones u otros que faciliten la organización). ➤ Permite el acceso a diferentes base de datos de diferentes proveedores como son Teradata 13.1, Oracle, SQL Server e Informix ➤ Permite realizar conexión a diferentes fuentes de información a través de ODBC y conectores nativos ➤ Permite el acceso a diferentes fuentes de datos de archivos planos. ➤ Permite el acceso a base de datos transaccionales ➤ Permite el acceso a base de datos no transaccionales ➤ Permite realizar perfilamiento de datos por tablas, campos o grupos de campos. ➤ Permite realizar perfilamiento al total de registros de las tablas, seleccionar un grupo o hacer una selección aleatoria. ➤ Administra usuarios y roles ➤ Administración de reglas ➤ Permite exportar los resultados del perfilamiento a archivos excel y PDF ➤ Permite exportar los resultados del proceso de reglas a archivos excel y PDF ➤ Consola de administración de resultados de los procesos (uso de recursos, status de ejecución) ➤ Permite crear servicios web a partir de las reglas u otras funcionalidades ➤ Las funciones de acceso a datos, perfilamiento, administrar usuarios y roles se diseñan mediante asistente de interfaz gráfico. ➤ Visualización de datos de múltiples fuentes sin usar scripts ➤ Capacidad para entregar los resultados en archivos planos y base de datos como MS SQL Server, Oracle, TeraData e Informix; luego de ejecutar el proceso de calidad aplicando las reglas. ➤ Establecimiento de métricas para evolución de la calidad de datos ➤ Realizar la creación de columnas virtuales. ➤ Permite crear reglas si necesidad de estar ligadas a un campo específico en un tabla o archivo plano. ➤ Capacidad para realizar un proceso de relacionamiento (maching) y éste debe permitir identificar datos coincidentes (mach), registros no coincidentes (no mach) y registro con probabilidad de coincidencia. ➤ Capacidad para realizar relacionamiento de tablas o archivos planos. ➤ Capacidad para permitir realizar relacionamiento de datos por más de un campo. 	4	0	3.4	3.3	3.8



	<ul style="list-style-type: none"> ➤ Capacidad para realizar relacionamiento de tablas por cualquier tipo de campo (numérico o texto) ➤ Transforma datos, luego de aplicar las reglas y generar data limpia ➤ Capacidad para llevar información estandarizada y limpia a fuentes de datos planos o base de datos transaccionales o no transaccionales (Oracle, Informix, Sql Server, Teradata) ya sean las mismas fuentes origen o una diferente. ➤ Capacidad para crear funcionalidad que permita ejecutar una alerta, esta puede ser en pantalla o correo ➤ Capacidad para integrar a las reglas de negocio códigos de lenguajes de programación como java u otro. ➤ Capacidad para trabajar múltiples usuarios ya sea development o de ejecución. ➤ Capacidad para manejar procesos de perfilamiento y ejecución de reglas de manera incremental u de reemplazo. ➤ Capacidad para realizar parseo de datos de tablas o archivos planos. ➤ Capacidad para manejar un repositorio de reglas para su reusabilidad. ➤ Navegación sobre todas las interfaces gráficas a través de los principales navegadores web: Internet Explorer (actualmente estándar en la SUNAT) y; Mozilla Firefox y Google Chrome (bajo aprobación de las áreas tecnológicas involucradas de la SUNAT), sin necesidad de instalación de software adicional en las estaciones clientes. ➤ Mostrar los resultados del perfilamiento y proceso de reglas de manera tabular y dashboards (gráficos). ➤ Capacidad para realizar seguimiento o trazabilidad del proceso de perfilamiento, estandarización y validación de datos de manera gráfica. 					
Exactitud	<ul style="list-style-type: none"> ➤ Verifica existencia de datos ➤ Verifica conformidad de datos ➤ Determina duplicidad de datos en campos ➤ Determina duplicidad de registros por tablas de base de datos y archivos planos ➤ Determina integridad de datos ➤ Identifica mach de dos archivos/tablas ➤ Identifica no mach de dos archivos/tablas ➤ Los resultados son de acuerdo a las reglas creadas ➤ Análisis de atributos máximos y mínimos ➤ Análisis de distribución de frecuencias 	3	0	3	3	3



	Interoperabilidad	<ul style="list-style-type: none"> ➤ Fácil integración con herramientas ETL ➤ El software tiene la capacidad crear reglas en servicios (web u otros) para ser usados por otras aplicaciones de SUNAT. ➤ Los servicios que genere el software de calidad de datos no deben limitarse a una sola plataforma para su funcionamiento. ➤ Integración con LDAP Directorio Activo de Windows ➤ Manejar reglas en múltiples proyectos, sin necesidad de replicarlas. 	3	0	3	2.5	2.5
	Seguridad	<ul style="list-style-type: none"> ➤ Gestión de privilegios por perfil y usuario. ➤ Integración con LDAP. ➤ El software facilita el trabajo de proyectos a múltiples usuarios. ➤ El software tiene la capacidad para limitar el acceso a la metadata. ➤ Gestionar acceso desde entorno web. ➤ Registro de auditoría como modificación y ejecución de los procesos. ➤ El software debe permitir consultar mediante interfaz gráfica (web o win), el registro de auditoría. ➤ No debe permitir almacenar de manera permanente datos en un repositorio intermedio. ➤ Capacidad para trabajar por perfiles de usuario. 	2	0	1.9	2	2
Confiabilidad	Nivel de Madurez	<ul style="list-style-type: none"> ➤ Permite monitorear fallas de procesos de calidad ➤ Permite hacer seguimiento a errores del proceso de calidad 	4	0	4	4	4
	Tolerancia a fallas	<ul style="list-style-type: none"> ➤ Los errores de un proceso de perfilamiento de datos no impide que el resto continúe operando ➤ El software soporta trabajar sin problemas con caracteres extraños y nulos. 	4	0	4	4	4
	Recuperación	<ul style="list-style-type: none"> ➤ Facilidad para identificar errores en los procesos ➤ Hacer seguimiento a los procesos de perfilamiento de datos que presentaron errores. ➤ Capacidad para volver a reprocesar los procesos que presentaron errores 	4	0	4	4	4
Usabilidad	Aprendizaje	<ul style="list-style-type: none"> ➤ Fácil uso de creación de nuevos proyectos o áreas de trabajo ➤ Fácil conexión a orígenes de datos ➤ Fácil uso de perfilamiento de datos ➤ Fácil creación y mantenimiento de reglas ➤ Facilidad para el uso en la interpretación de resultados ➤ Facilidad para ver el detalle de los datos ➤ Facilidad para crear indicadores como metas de seguimiento de calidad de datos 	3	0	2.7	2.8	2.9



	Comprensibilidad	<ul style="list-style-type: none"> ➤ El usuario entiende la interfaz de fuentes de datos ➤ El usuario entiende interfaz de perfilamiento ➤ El usuario entiende la interfaz de crear reglas de alto nivel ➤ El usuario entiende integrar reglas con campos de fuentes de datos ➤ El usuario entiende resultados estadísticos emitidos por el sistema ➤ El usuario entiende la gestión de usuarios. 	3	0	2.3	2.6	2.3
	Operabilidad	<ul style="list-style-type: none"> ➤ El usuario puede hacer uso del acceso a datos ➤ El usuario puede hacer uso de la funcionalidad de perfilamiento ➤ El usuario puede hacer uso del manejo de seguridad ➤ El usuario puede hacer uso del manejo de reglas ➤ El usuario puede acceder a los resultados del proceso de perfilamiento y ejecución de reglas ➤ El usuario puede crear servicios web (http, SOAP, REST, RMI) ➤ El usuario puede publicar servicios web (http, SOAP, REST, RMI) ➤ El usuario puede acceder a los servicios web 	3	0	3	3	2.8
	Atractividad	<ul style="list-style-type: none"> ➤ Tiene interfaz de conexión a las fuentes de datos. ➤ Tiene interfaz de perfilamiento de datos. ➤ Tiene interfaz de gestión de reglas. ➤ Tiene interfaz de los resultados presentados por el proceso de perfilamiento. ➤ Tiene interfaz del proceso de ejecución de reglas. ➤ El sistema muestra mensajes durante el tiempo de procesamiento. ➤ Los colores de las pantallas son intensos, provocando que se cansa fácilmente la vista del usuario. 	3	0	3	3	3
Eficiencia	Comportamiento con respecto al Tiempo	<ul style="list-style-type: none"> ➤ El sistema no se muestra lento ➤ Los detalles de los datos que se desean consultar se visualizan de manera inmediata 	4	0	4	4	4
	Comportamiento con respecto a Recursos	<ul style="list-style-type: none"> ➤ La herramienta permite optimizar el uso de memoria para su procesamiento ➤ La herramienta permite optimizar el uso de acceso a disco duro ➤ La herramienta durante el procesamiento de los trabajos provoca que otras aplicaciones no se vean impactados en su rendimiento 	4	0	3	3.3	3.3
	Estabilidad	<ul style="list-style-type: none"> ➤ Facilidad para hacer upgrades 	4	0	3	3	3
Portabilidad	Adaptabilidad	<ul style="list-style-type: none"> ➤ El software corre S.O. como UNIX/LINUX ➤ El software se podrá ejecutar en otra arquitectura de hardware a la instalada inicialmente ➤ Soporta una arquitectura orientada a servicios, es decir, que los procesos de calidad de datos puedan convertirse en servicios y ser así invocados desde los distintos aplicativos de EL CLIENTE. 	3	0	3	3	3
	Facilidad de Instalación	<ul style="list-style-type: none"> ➤ El software podrá ser instalado por personal de SUNAT 	3	0	3	3	3



	Capacidad de reemplazo	<ul style="list-style-type: none"> ➤ El software de calidad de datos podrá ser sustituido otro de manera similar. 	3	0	3	3	3
	Co-existencia	<ul style="list-style-type: none"> ➤ El software puede funcionar con otras aplicaciones de otros proveedores ➤ El software puede funcionar con aplicaciones realizadas por SUNAT. 	3	0	3	3	3
CALIDA EN USO			40	0	39	39.8	39.1
Efectividad	Acceso a datos	<ul style="list-style-type: none"> ➤ Permite el acceso a diferentes base de datos de diferentes proveedores como son Teradata 13.1, Oracle, SQL Server e Informix ➤ Permite realizar conexión a diferentes fuentes de información a través de ODBC y conectores nativos ➤ Permite el acceso a diferentes fuentes de datos de archivos planos (txt, CSV, xml, etc.) ➤ Permite el acceso a base de datos transaccionales ➤ Permite el acceso a base de datos no transaccionales. 	2	0	1.9	2	2
	Nivel de perfilamiento	<ul style="list-style-type: none"> ➤ Crear perfilamiento por cualquier campo y cualquier tipo de dato. ➤ Crear perfilamiento de datos por tablas, campos o grupos de campos. ➤ Procesamiento paralelo de perfilamiento ➤ Existe consola de procesamiento para perfilamiento de datos ➤ Permite procesar trabajos de perfilamiento planificados ➤ Permite procesar trabajos de perfilamiento a demanda 	2	0	2	2	2
	Nivel de validación de datos	<ul style="list-style-type: none"> ➤ Determina existencia de datos ➤ Determina conformidad de datos ➤ Determina duplicidad de datos ➤ Determina integridad de datos ➤ Identifica mach de dos archivos/tablas ➤ Identifica no mach de dos archivos/tablas ➤ Identifica registros con probabilidad de coincidencia 	2	0	2	2	2
	Manejo de reglas	<ul style="list-style-type: none"> ➤ Permite analizar valores de campos como direcciones, nombres de personas naturales o jurídicas. ➤ Permite realizar un proceso de relacionamiento (maching) y éste debe permitir identificar datos coincidentes (mach), registros no coincidentes (no mach) y registro con probabilidad de coincidencia. ➤ Maneja consola de creación de reglas de manera gráfica (sin escribir código). 	2	0	1.8	2	1.8
	Presentación de resultados	<ul style="list-style-type: none"> ➤ Los resultados son de acuerdo a las reglas creadas ➤ Muestra estadísticas de manera tabular y gráfica 	2	0	2	2	2
	Productividad	Tiempo para crear proyectos	<ul style="list-style-type: none"> ➤ Para acceso a las fuentes de datos no se necesita conocimiento de base de datos. ➤ La creación de perfilamiento será creado por personal del negocio. ➤ La creación de perfilamiento será creado por personal de TI. ➤ No se necesitan conocimientos de SQL para realizar las reglas. ➤ Para la lectura de los resultados no se necesitan conocimientos avanzados de estadísticas. 	10	0	9.5	10



Seguridad	Manejo de usuarios y roles	<ul style="list-style-type: none"> ➤ El software limita el acceso a los diferentes componentes de un proyecto de acuerdo a los perfiles del usuario. ➤ El software se integra con LDAP. ➤ El usuario podrá acceder a la metadata. 	5	0	5	5	5
	Backup	<ul style="list-style-type: none"> ➤ El software permite realizar backups e integrarlo al proceso existente de SUNAT. 	5	0	5	5	5
Satisfacción	Mensajes de procesamiento	<ul style="list-style-type: none"> ➤ El software muestra mensajes de procesamiento de tareas para tareas de perfilamiento. ➤ El software muestra mensajes de procesamiento de procesamiento de reglas. 	2.5	0	2.5	2.5	2.5
	Resultados	<ul style="list-style-type: none"> ➤ Los resultados que presenta el software se pueden visualizar en pantalla ➤ Los resultados que presenta el software se pueden exportar a Excel y pdf. ➤ EL software muestra resultados de perfilamiento de datos ➤ El software muestra resultados de procesos de calidad de datos 	2.5	0	2.4	2.4	2.5
	Navegación	<ul style="list-style-type: none"> ➤ El software demuestra se intuitivo en su navegación por sus diferente opciones. ➤ El software cuenta con ayuda en línea para todas las opciones que presenta 	2.5	0	2.5	2.5	2.5
	Cientes de acceso	<ul style="list-style-type: none"> ➤ El software puede ser utilizado a través de un cliente desktop, en Windows 7/8/XP. ➤ El software puede ser utilizado a través de un cliente web, con Internet explorer 7.0 o superior y Mozilla Firefox 12 o superior. ➤ Utiliza un entorno IDE para desarrollo. 	2.5	0	2.4	2.4	2.3

8. ANÁLISIS COMPARATIVO DE COSTO-BENEFICIO

COSTOS:

Licenciamiento: Se estima una plataforma para hacer Calidad de Datos con 18 licencias para usuarios administradores y analistas de Calidad de datos para los ambientes de producción y desarrollo – pruebas.

El licenciamiento aproximado del Software Informática Data Quality para los ambientes de producción y desarrollo – pruebas tiene un costo de \$ 630,120.00 (incluye IGV).

El licenciamiento aproximado de Trillium Software System para los ambientes de producción y desarrollo – pruebas tiene un costo de \$ 342,200.00 (incluye IGV).

El licenciamiento aproximado de InfoSphere Data Quality para los ambientes de producción y desarrollo – pruebas tiene un costo de \$ 1'745,601.75 (incluye IGV).

El soporte y mantenimiento: en caso Trillium Software System tiene un costo de \$ 205,3200.00 (incluye IGV), Informática Data Quality tiene un costo de \$ 415,879.20 (incluye IGV) y para InfoSphere Data Quality tiene un costo de \$ 718,951.23 (incluye IGV) en todos los casos es por 3 años.



Capacitación: En el caso de Informática Data Quality el costo de capacitación es de \$32,450.00 (incluye IGV) aproximadamente, en el caso de Trillium Software System la capacitación tendría un costo de \$ 9,440.00 (incluye IGV) y para InfoSphere Data Quality \$ 25,960.00 (incluye IGV)

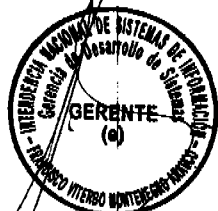
BENEFICIOS:

Los beneficios esperados se verán reflejados en brindar mejores servicios a los contribuyentes y ser más efectivos en el control del cumplimiento, ya que se tendrá un soporte de datos para generar información fiable y se contribuirá a mejorar la productividad de los trabajadores en los procesos del negocio y atención al contribuyente. Así mismo se espera identificar nuevas oportunidades para la SUNAT, implementación de nuevos productos o servicios que aumenten la recaudación, reducción de plazos de entrega de información a usuarios internos o externos, lo que redundará en la mejora de la imagen institucional y reducción de riesgos por decisiones tomadas sin sustento basado en hechos o datos.

9. CONCLUSIÓN

Las tres herramientas analizadas cumplen con los requisitos técnicos mínimos, por lo tanto se recomienda realizar un proceso por concurso en el que participen estas herramientas además de cualquier otra que satisfaga los requerimientos técnicos mínimos.

10. FIRMA



YURY HAROLD CASTAÑEDA PEDROZA
Jefe Div. de Desarrollo de Sist. Analíticos (e)
GERENCIA DE DESARROLLO DE SISTEMAS