Table of Contents

1. PraaS Overview  

2. The Profiling Goal  

3. What do you get from Profiling?  

4. How PraaS Improves the Profiling Experience  

5. What is the Profiling Process?  

6. PraaS Functional Components  

7. PraaS Planning and Administration  

GlobalIDs.com
1. **PraaS Overview**

Global IDs PraaS is a scalable enterprise-level profiling solution for large data landscapes. It utilizes an agent-based, massively parallel, distributed computing architecture aimed at building transparency and semantic understanding of the data landscape.

2. **The Profiling Goal**

**Why profile?**

*Find out what you actually have in files and documents.*

- This includes finding “bad” data, and finding sensitive data where it should not be.
- Knowing “field names” is a start to understanding the data, but there is no guarantee that the names really mean what you may assume. This is more than a Data Quality issue.
- How many different kinds of “cost” fields can you have? Find telephone numbers in a person’s First Name field? Confidential identifiers or financial information in a Description? Does R&D speak the same language as Sales?
- And if you don’t look past the wrapper, then you are at risk in any development or deployment that uses the data.

3. **What do you get from Profiling?**

The ultimate profiling deliverable is a definition and set of rules that can be used for driving data quality for the columns or fields in the applications systems.

**The starting point:** Basic profiling statistics measure the distribution of the values across all the records in a table. These statistics may be by the value itself, by pattern, by clusters of values, and other characteristics of the values found.

The next step is to assign meaning, and establish rules for which values are valid and which are not. This is historically done by analysts. However, Global IDs uses statistical analysis and “domain knowledge” to suggest what is in the fields.

1. For values that cannot be matched to domains, Global IDs identifies statistical outliers for honing the data quality rules, and can compare “like” fields.

2. On the “domain knowledge” side, Global IDs has an encyclopedia of standard codes and values from governmental, industry, and standards organizations like ISO to suggest what the data is. For example, a field may contain ISO 2-character Country Codes in over 90% of the records. Global IDs can do this because it is pre-loaded with those standards. Global IDs goes a step further to allow detecting identifiers and more ordinary fields like first names and telephone numbers.
4. How PraaS Improves the Profiling Experience

PraaS incorporates three technical steps forward, each with attendant advantages:

1. The ability to deploy multiple “Federated” instances, providing:
   - Project or organizational autonomy for respective projects on one hand
   - The ability to share Metadata among the federate GIDs instances.

2. Resource Scalability & Performance by GIDs instances in a Cloud environment, providing:
   - Scaling the resources, taking advantage of parallelism, to meet processing demands

3. An all-Web interface to eliminate desktop software management (All you need is the browser.)

4. A simplified interface and functional capability to optimize manual effort, both in profiling and managing the profiling process

5. What is the Profiling Process?

At this time, PraaS profiles data in networked relational database systems like Oracle, SQL Server, MySQL, and DB2. For these databases, the profiling process has four logical steps:

1. Define the Connection (analogous to a “Data Source” in Microsoft’s ODBC).
2. Get the Data Dictionary information from the connected DBMS instance.
3. Profile the columns of interest.
4. Clarify the meanings of the values found.

The first three steps usually provide data analysts the information they need to perform the fourth step. As noted, Global IDs also jump-starts the fourth step, allowing the analyst to skip the definition and simply go to verification for many fields.

6. Better, Cheaper, Faster: Automating the Process

Except for getting the connections and authorizations, the first three steps can be largely consolidated into a “one touch” approach, where you enter the connection information – or just select an already identified connection - and press a “Profile” button.

There is one practical impediment to this scenario: You need to avoid resource contention between the profiling process and the “production” application for some high-volume tables.
• This may require holding back some columns from the “one touch” job. This in-turn requires a capability for managing the subsequent profiling of these columns. While not included in the initial PraaS offering, improvements to get to a “one touch” capability are in process.

7. PraaS Functional Components

PraaS has two parts.

The PraaS Client Portal is where the data analysts or scientists profile the data. The Administration component is broken out among several portals. These portals manage the federated server hierarchies and the cloud, and access control for the respective PraaS Client portals. These are covered in separate Guides.

8. PraaS Planning and Administration

1. The Data Landscape
2. The Data Governance Landscape
3. Setup and Administration
   • GIDs Federated Systems Deployment & Administration
   • GIDs Foundation Administration
     i. Access & Security
        • The Global IDs Administration takes a sophisticated approach when dealing with access control. Every user is compartmentalized in each parent or child server. It is then from each of those servers have their own users and permissions. In addition, each user can be transferred from the parent server to the child server or from the child server to the parent server. Global IDs also supports Kerberos and LDAP protocol, allowing administrators to add users from both protocols.
     ii. Workflow Setup
     iii. Ongoing Monitoring
        • In addition to data profiling, Global IDs Profiling as a Service also provides monitoring, allowing you to keep the data landscape updated.
   • The Cloud
   • The Process
## Global IDs Product Suite vs. PraaS

<table>
<thead>
<tr>
<th></th>
<th>GLOBAL IDs PRODUCT SUITE</th>
<th>PROFILING AS A SERVICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deployment</td>
<td>Deployed at customer site</td>
<td>Managed by Global IDs, no customer-side deployment necessary</td>
</tr>
<tr>
<td>Profiling</td>
<td>Provides extensive profiling options</td>
<td>Allows One-Touch Profiling &amp; advanced options for power-users</td>
</tr>
<tr>
<td>User Interface</td>
<td>Desktop interface and web interface</td>
<td>Web interface for easy access, use, and reporting</td>
</tr>
<tr>
<td>Backend</td>
<td>Localized Habitat creation</td>
<td>Federated instance and resource allocation</td>
</tr>
<tr>
<td>Networking</td>
<td>Moderate customer-side Network Manipulation</td>
<td>Minimal customer-side network manipulation; only needs ports and IP of networked database</td>
</tr>
</tbody>
</table>

### Questions?

Contact us to schedule a demo or to get started.

Toll Free:  
[888] 514-0192  
GlobalIDs.com  • info@GlobalIDs.com

Local:  
[609] 683-1066  
[646] 201-9498