



Tendances Logicielles

Présentation EAS - GNR

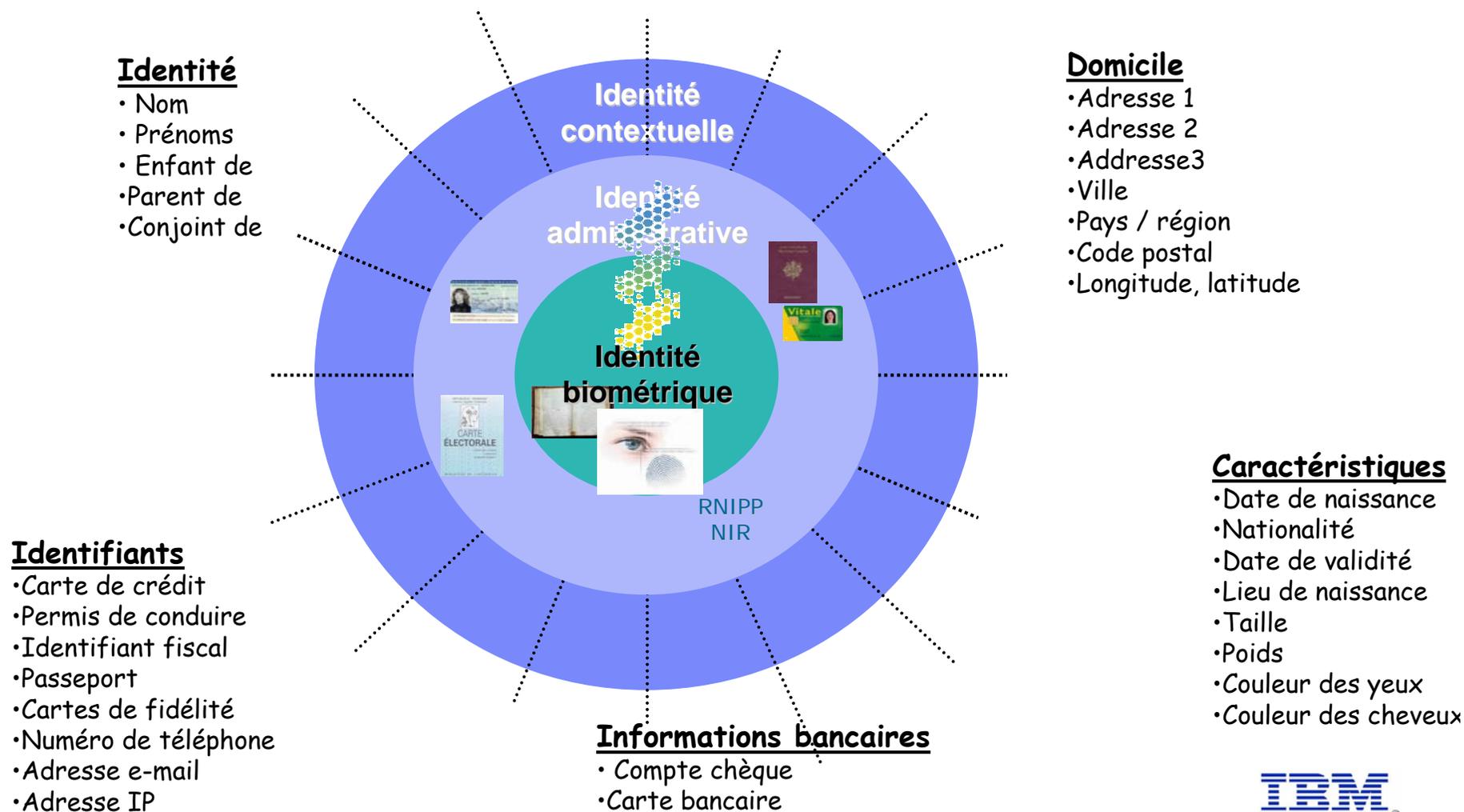
David Decloux
Information On Demand Architect
IBM Software Group



ON DEMAND BUSINESS™

© 2008 IBM Corporation

L'identité du citoyen est une combinaison de données personnelles différentes : une notion complexe et aux multiples facettes



Nécessité d'accumuler automatiquement et de traiter en mettant en liaison toutes les informations internes et externes

La Fraude à l'identité : Une préoccupation croissante,

Blanchiment d'argent : "une "industrie" de 750 à 1750 Mds \$ (2 à 5 % du PIB) qui croît plus vite que l'économie mondiale
FMI

Fausse identité coûtent environ 30 millions d'Euros à Air France

La sécurité sociale évalue la fraude à 1% de son budget soit 1,4 milliard d'Euros

90 % des amendes impayées de la SNCF sont dues à des fausses identités (202 millions d'Euros)

Aux Etats Unis, 27 millions d'américains (un peu moins de 10% de la population totale) ont été victimes d'une usurpation d'identité lors des cinq dernières années

Les fraudes dues à la collusion entre les employés et des tiers représentent 48% de l'ensemble des fraudes .
American Banker

Les pertes liées à l'usurpation d'identité sont en croissance de 300 % (estimation : 2 Mds \$ fin 2005)
Aberdeen Group

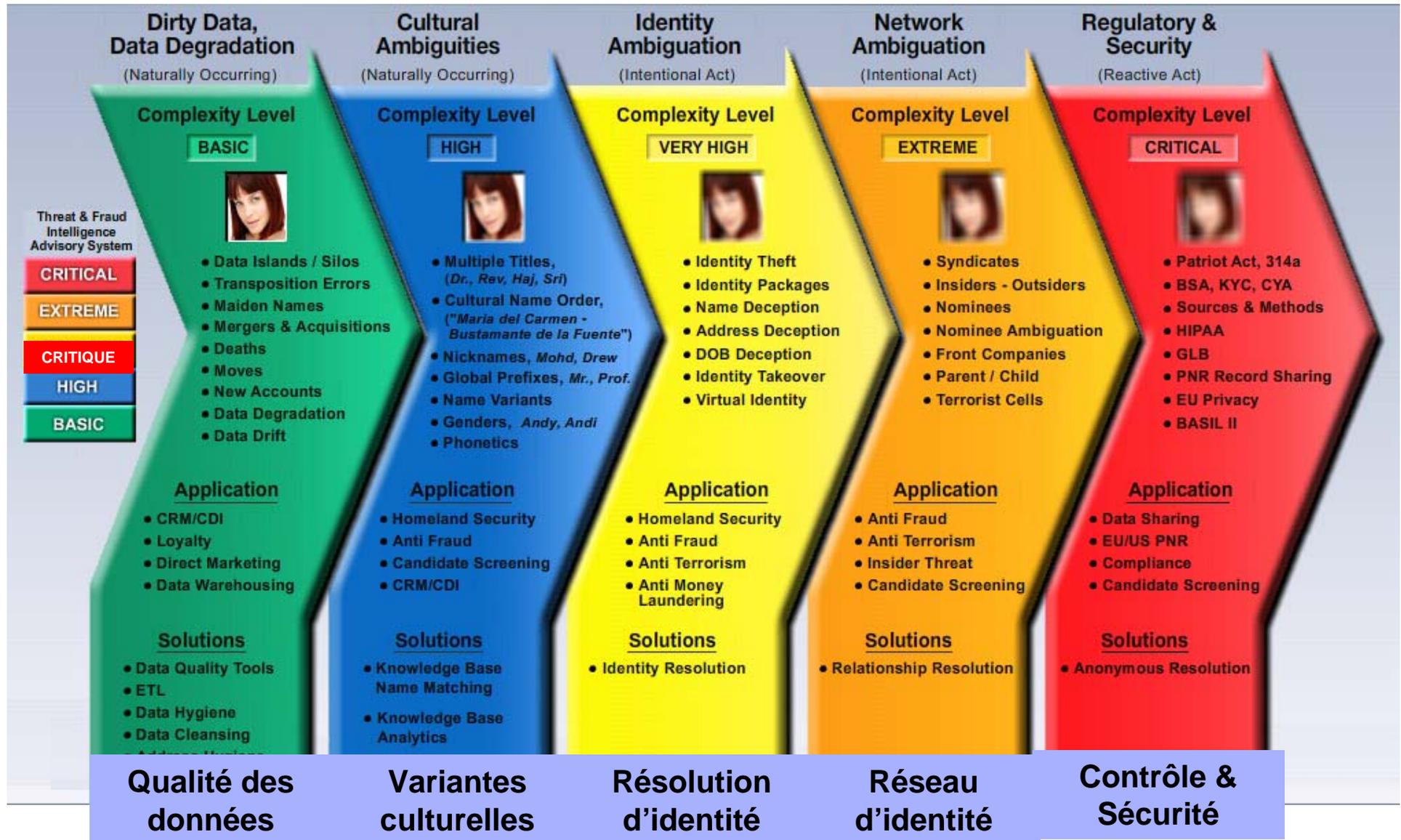
Toutes les 79 secondes, aux Etats-Unis, une identité est usurpée, un compte bancaire sous fausse identité est ouvert, des achats frauduleux sont effectués

Taux de croissance annuel des fraudes à l'identité : 42% aux USA, 41% en GB, et probablement 40% au moins en France.

En 2002, les fraudes à l'identité enregistrées par la police ont causé un préjudice dans le monde de 73,8 milliards de dollars (63 milliards d'Euros)



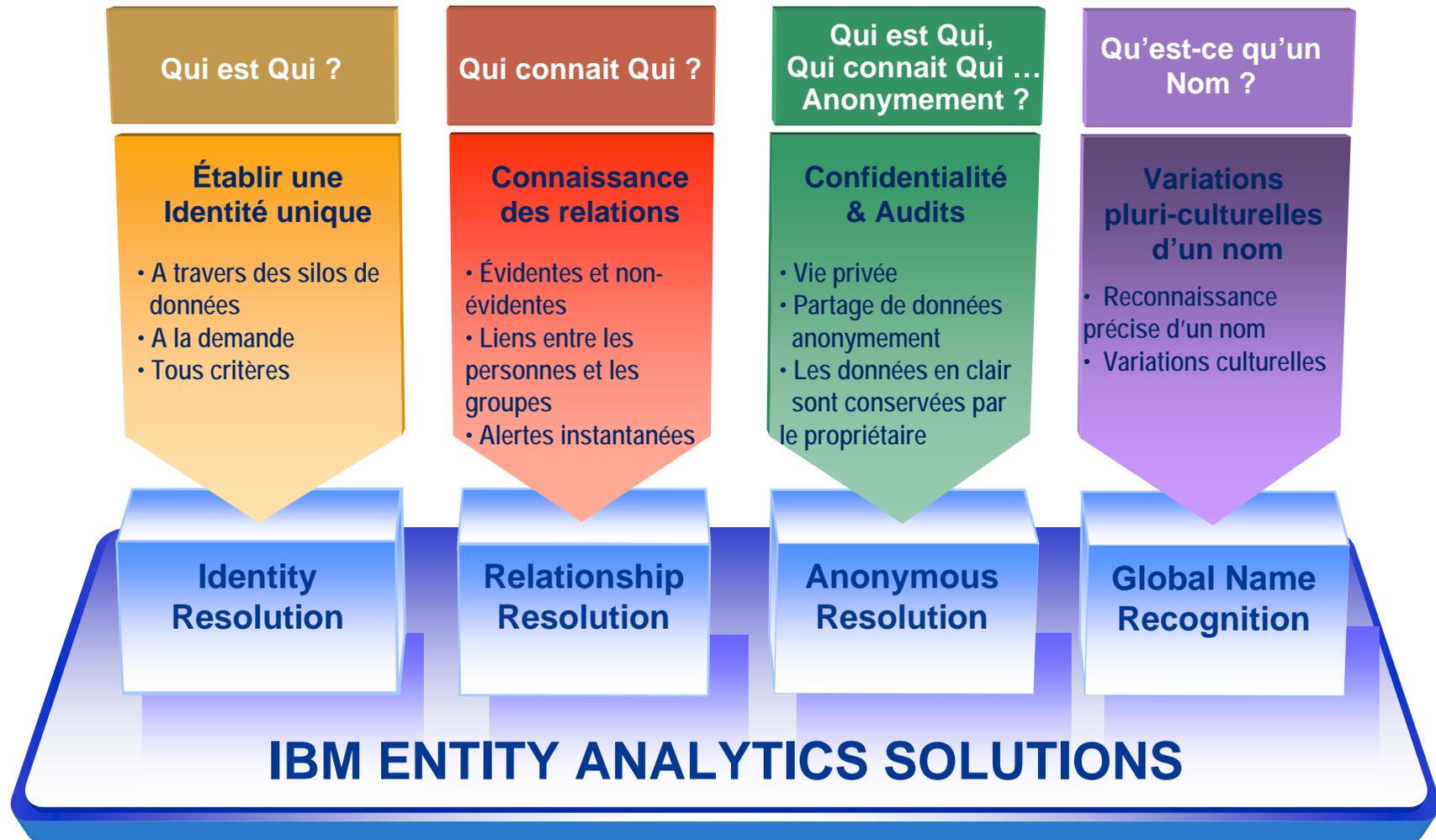
Les "5" niveaux relatif aux erreurs et à l'ambiguïté de l'identité



Agenda

- EAS – GNR : une offre modulaire
- Entity Analytics Solution
 - Identity Resolution
 - Relationship Resolution
 - Anonymous resolution
- Global Name Recognition

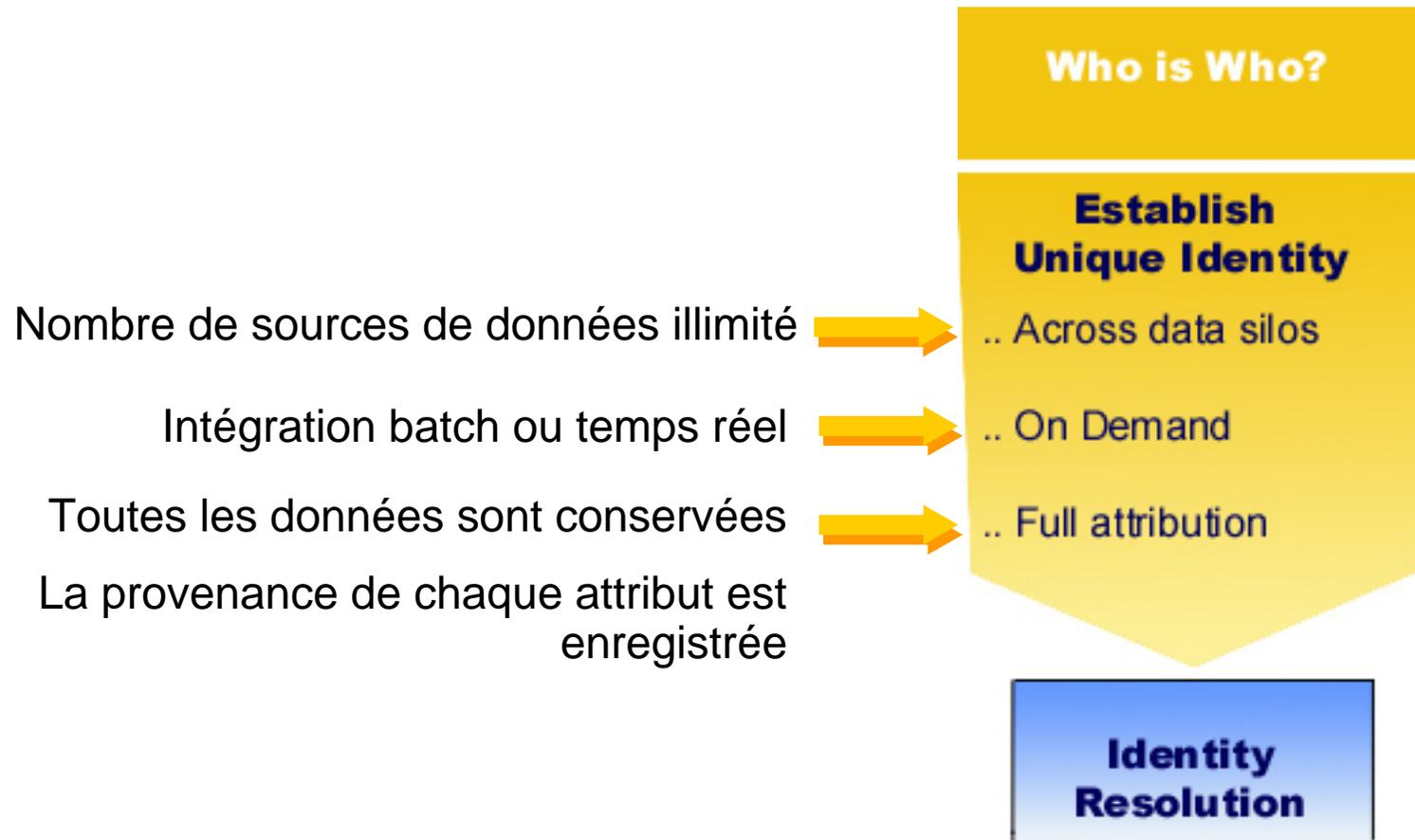
EAS et GNR proposent 4 fonctions



Agenda

- EAS – GNR : une offre modulaire
- Entity Analytics Solution
 - Identity Resolution
 - Relationship Resolution
 - Anonymous resolution
- Global Name Recognition

Identity Resolution



Résolution d'identité

“Qui est qui?” “Mauvaise représentation des attributs”

Nom: surnom, abréviation, consonances...

(Kate Green > **C**ate **G**reene)

Résidence: numéro, type de rue, nom...

(25 North Bourne St > **52 West Born Ave**)

Numéro d'identité: NSS, Passeport, Permis de conduire, numéro INE...

(026-65-3632 > 0**62**-65-36**23**)

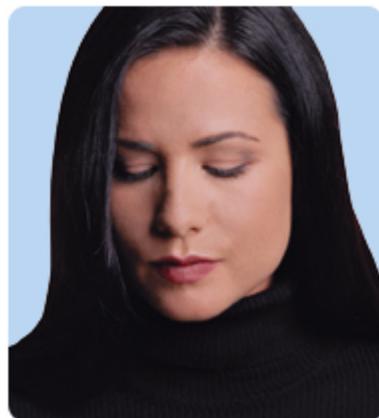
Date de naissance:

(07-08-1964 > 0**8**-0**7**-1964)

Résolution d'identités



Ce module permet de reconnaître une personne physique ou morale utilisant de multiples identités



Dr Katrin Dupont
 1 Ave Dumesnil
 78230 Saint-Cloud
 Tel :01 49 05 50 84
 CDF : 2640809737615
 DDN :07/08/64
 Tel : 068588345



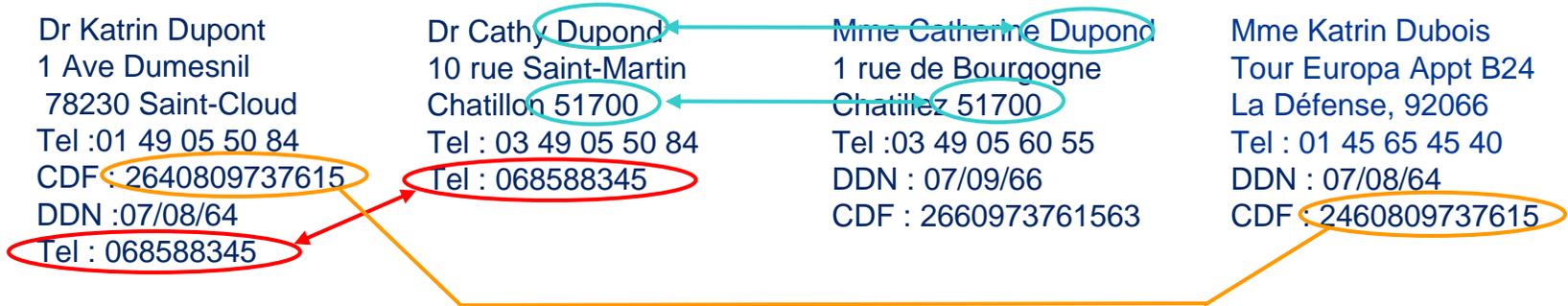
Dr Cathy Dupond
 10 rue Saint-Martin
 Chatillon 51700
 Tel : 03 49 05 50 84
 Tel : 068588345



Mme Catherine Dupond
 1 rue de Bourgogne
 Chatillez 51700
 Tel :03 49 05 60 55
 DDN : 07/09/66
 CDF : 2660973761563



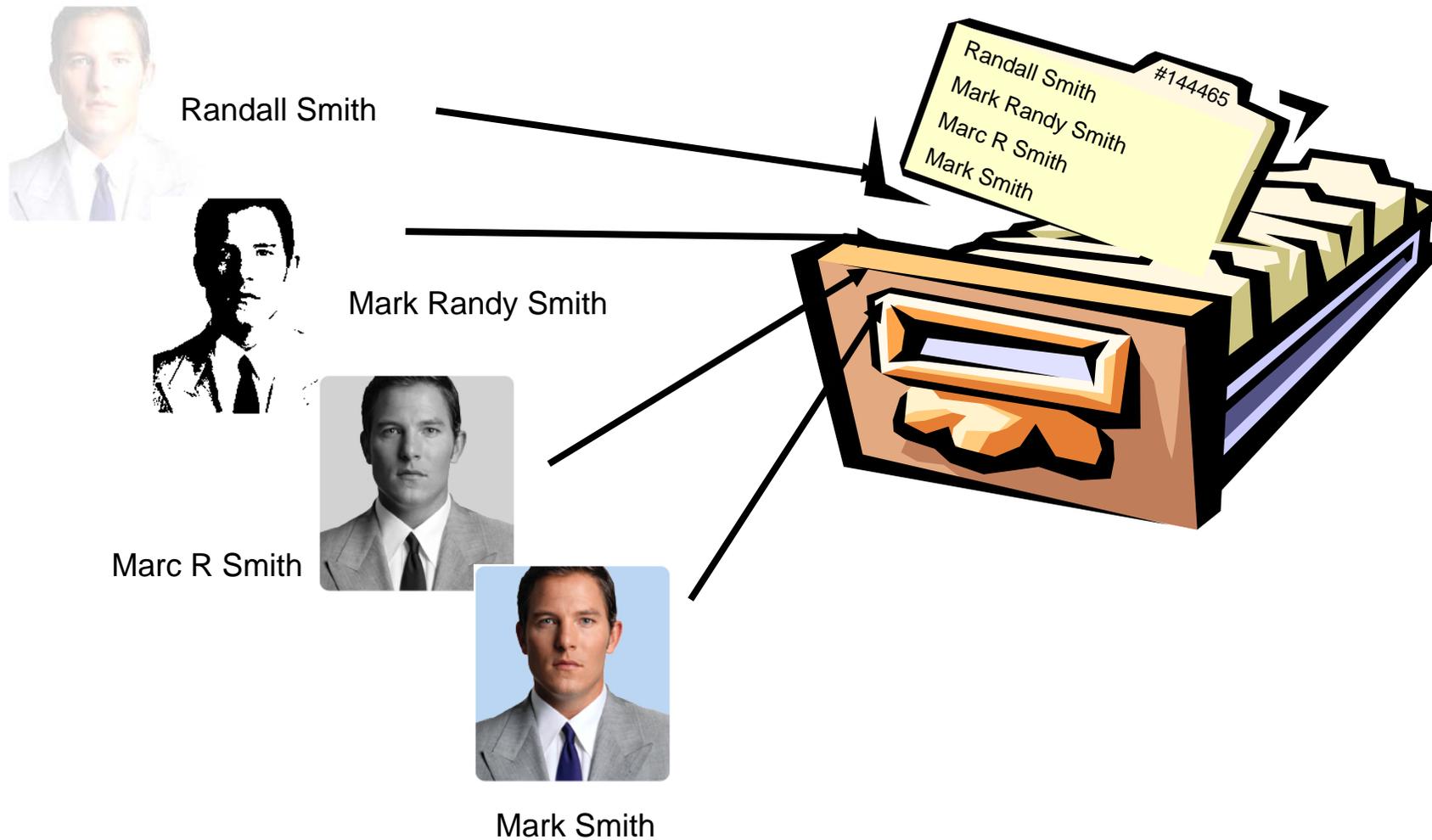
Mme Katrin Dubois
 Tour Europa Appt B24
 La Défense, 92066
 Tel : 01 45 65 45 40
 DDN : 07/08/64
 CDF : 2460809737615



Sur la base d'informations pré-définies ou extensibles

Nom	Adresse	Identifiants	Attributs
<ul style="list-style-type: none">– Nom de famille– Prénom– Surnom– Nom d'usage– Nom d'organisation– Alias– Etc...	<ul style="list-style-type: none">– Adresse 1– Adresse 2– Adresse 3– Ville– Etat/Province– Code Postal– Pays– Etc...	<ul style="list-style-type: none">– Carte de crédit– Permis de conduire– Compte bancaire– Numéro CNI– Numéro Passeport– Téléphone– Adresse Courriel– NIR– SIREN– ... autres (définis à la demande)	<ul style="list-style-type: none">– Date naissance– Nationalité– Lieu de Résidence– Lieu de naissance– Taille– ... autres (définis à la demande)

Les entités sont stockées dans le référentiel EAS



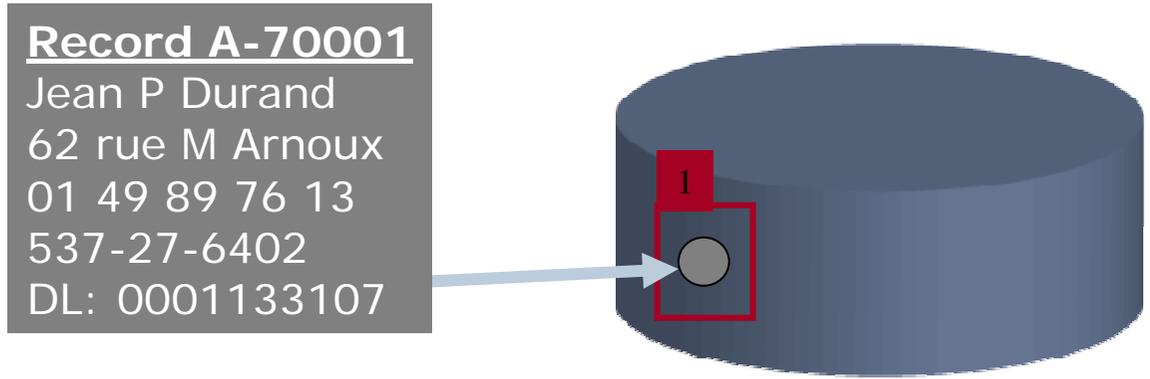
Exemple d'entités enregistrées sous une identité unique

Pointeurs vers les systèmes source



Entity #144465		
Names	Marc R Smith	A-#70001
	Randal Smith	B-#009102
	Mark Randy Smith	C-#6251
Address	123 Main St.	A-#70001
	456 First Street	C-#6251
Phones	(713) 730-5769	A-#70001
	(713) 731-5577	B-#009102
	(713) 731-5577	C-#6251
SSN	537-27-6402	A-#70001
DL	1133107	A-#70001
	1133107	C-#6251
DOB	06/17/1974	B-#009103

Injection d'un premier enregistrement Que se passe-t-il ?



EAS ID #144465

Noms	Jean P Durand	A-#70001
Adresse	62 rue M Arnoux	A-#70001
Tél	01 49 89 76 13	A-#70001
CNI	537-27-6402	A-#70001
PC	1133107	A-#70001



Une nouvelle entité (#1) est créée dans la base



Injection d'un deuxième enregistrement

Record A-70001

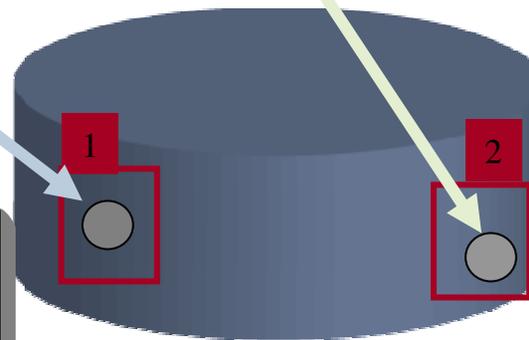
Jean P Durand
 62 rue M Arnoux
 01 49 89 76 13
 537-27-6402
 PC: 0001133107

Record B-00009103

Pierre J Durand
 DDN: 17/06/1934
 01 31 33 22 09

EAS ID #144465

Noms	Jean P Durand	A-#70001
Adresse	62 rue M Arnoux	A-#70001
Tél	01 49 89 76 13	A-#70001
CNI	537-27-6402	A-#70001
PC	1133107	A-#70001



EAS ID #321489

Noms	Pierre J Durand	B-#9103
DDN	17/06/1934	B-#9103
Tél	01 31 33 22 09	B-#9103



Pas de corrélation pertinente entre les identités –
 création d'une deuxième entité





Injection d'un troisième enregistrement

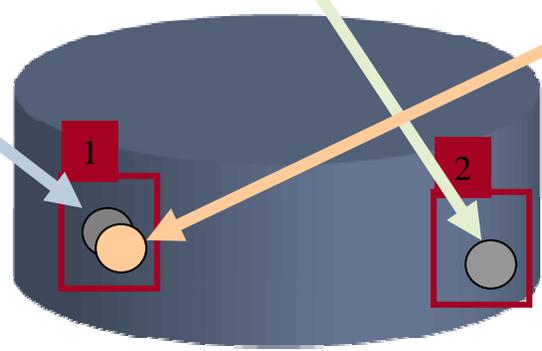
Record A-70001
 Jean P Durand
 62 rue M Arnoux
 01 49 89 76 13
 537-27-6402
 PC: 0001133107

Record B-00009103
 Pierre J Durand
 DDN: 17/06/1934
 01 31 33 22 09

Record C-6251
 Jean-Pierre Durand
 10 rue Sadi Carnot
 01 31 33 22 09
 PC: 1133107

EAS ID #144465

Noms	Jean P Durand	A-#70001
	Jean-Pierre Durand	C-#6251
Adresses	62 rue M Arnoux	A-#70001
	10 rue Sadi Carnot	C-#6251
Tél	01 49 89 76 13	A-#70001
	01 31 33 22 09	C-#6251
CNI	537-27-6402	A-#70001
PC	1133107	C-#6251
	1133107	A-#70001



EAS ID #321489

Noms	Pierre J Durand	B-#9103
DDN	17/06/1934	B-#9103
Tél	01 31 33 22 09	B-#9103

Des attributs sont rapprochés avec l'entité #1 = "Résolution"





Contexte à l'injection – 3ème enregistrement

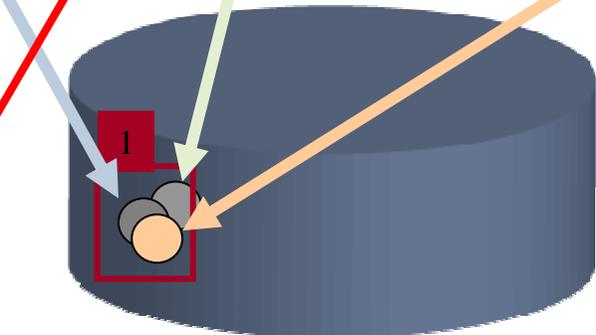
Record A-70001
 Jean P Durand
 62 rue M Arnoux
 01 49 89 76 13
 537-27-6402
 PC: 0001133107

Record B-00009103
 Pierre J Durand
 DDN: 17/06/1934
 01 31 33 22 09

Record C-6251
 Jean-Pierre Durand
 10 rue Sadi Carnot
 01 31 33 22 09
 PC: 1133107

EAS ID #144465

Noms	Jean P Durand	A-#70001
	Jean-Pierre Durand	C-#6251
	Pierre J Durand	B-#9103
Adresses	62 rue M Arnoux	A-#70001
	10 rue Sadi Carnot	C-#6251
Tél	01 49 89 76 13	A-#70001
	01 31 33 22 09	B-#9103
	01 31 33 22 09	C-#6251
CNI	537-27-6402	A-#70001
PC	1133107	C-#6251
	1133107	A-#70001
DDN	17/06/1934	B-#9103



L'identité #2 correspond à l'identité #1 –
 Les entités sont jointes



Injection d'un quatrième enregistrement

Record A-70001
 Jean P Durand
 62 rue M Arnoux
 01 49 89 76 13
 537-27-6402
 PC: 0001133107

Record B-00009103

Pierre J Durand
 DDN: 17/06/1934
 01 31 33 22 09

Record C-6251

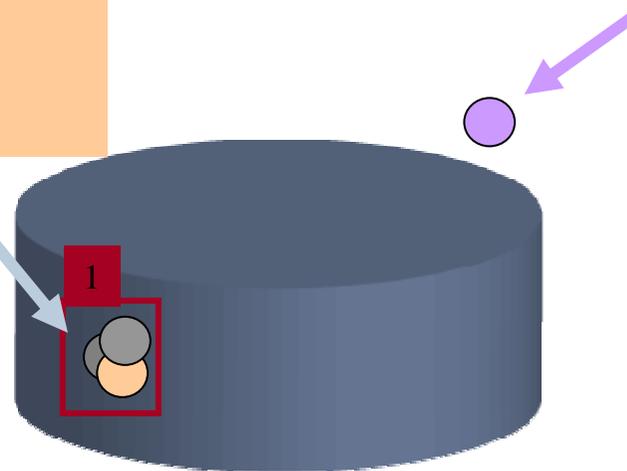
Jean-Pierre Durand
 10 rue Sadi Carnot
 01 31 33 22 09
 PC: 1133107

Record D-7700214

Cpt Pierre Durand
 DDN: 17/06/1934
 01 31 33 22 09
 423-22-7027

EAS ID #144465

Noms	Jean P Durand	A-#70001
	Jean-Pierre Durand	C-#6251
	Pierre J Durand	B-#9103
Adresses	62 rue M Arnoux	A-#70001
	10 rue Sadi Carnot	C-#6251
Tél	01 49 89 76 13	A-#70001
	01 31 33 22 09	B-#9103
	01 31 33 22 09	C-#6251
CNI	537-27-6402	A-#70001
PC	1133107	C-#6251
	1133107	A-#70001
DDN	17/06/1934	B-#9103



L'enregistrement est rapproché de l'entité #1



Injection d'un quatrième enregistrement

Record A-70001

Jean P Durand
62 rue M Arnoux
01 49 89 76 13
537-27-6402
PC: 1133107

Record C-6251

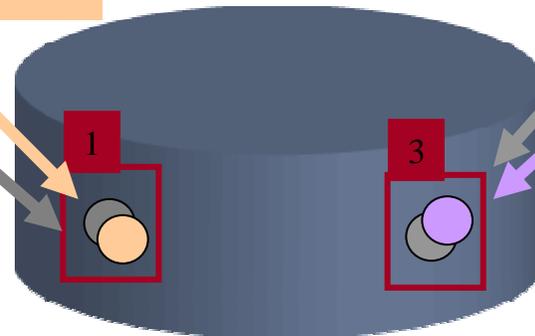
Jean-Pierre Durand
10 rue Sadi Carnot
01 31 33 22 09
PC: 1133107

Record D-7700214

Cpt Pierre Durand
17/06/1934

Record B-00009103

Pierre J Durand
DDN: 17/06/1934
01 31 33 22 09



Un conflit est identifié –
le contexte est corrigé
= "Découplage"

EAS ID #144465

Noms	Jean P Durand	A-#70001
	Jean-Pierre Durand	C-#6251
Adresses	62 rue M Arnoux	A-#70001
	10 rue Sadi Carnot	C-#6251
Tél	01 49 89 76 13	A-#70001
	01 31 33 22 09	C-#6251
CNI	537-27-6402	A-#70001
PC	1133107	C-#6251
	1133107	A-#70001



EAS ID #321463

Noms	Cpt Pierre Durand	D-#7700214
	Pierre J Durand	B-#9103
Tél	01 31 33 22 09	D-#7700214
	01 31 33 22 09	B-#9103
CNI	423-22-7027	D-#7700214
DDN	17/06/1934	D-#7700214
	17/06/1934	B-#9103



Fonctionnement résumé

Injection continue et analyse en temps réel ou par fichiers

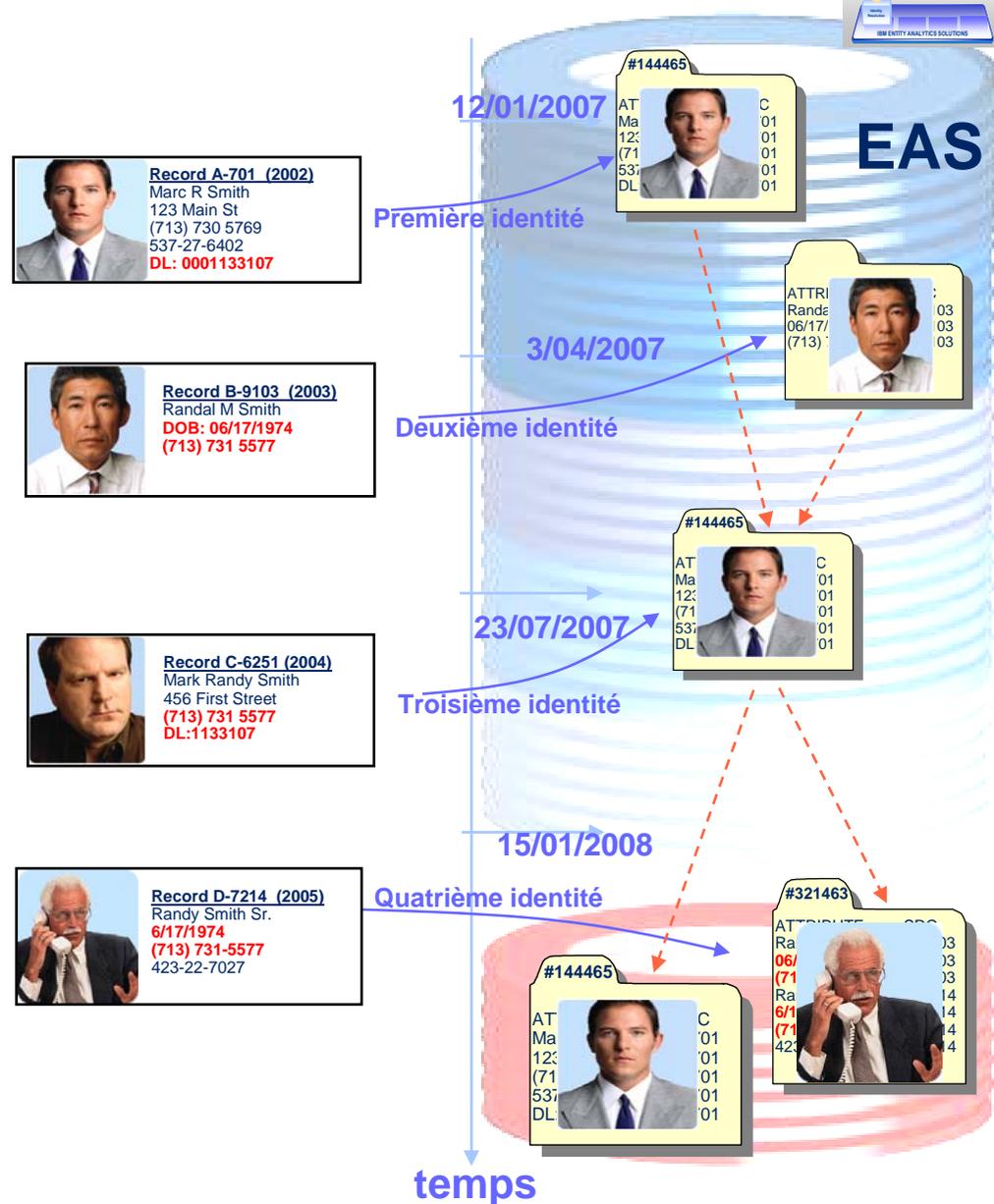
Tous les attributs sont utilisés

Le contexte est persistant

Pas d'apprentissage préalable de l'outil

Auto-correctif et indépendant des séquences

Fonctionnement viable sur le long terme



Exemple 1: nom et adresse proches, date de naissance et numéro de RIB identiques

The screenshot displays the IBM Relationship Resolution interface. The main window, titled "Graphique de l'alerte", shows a network diagram of relationships. Two central nodes represent individuals: "GLENN PERTERSON START" (ID 47793) and "GLEN PETERSON STUART" (ID 30267). They are connected to several other nodes, including addresses like "4 Rue Du Hammer 20" and "4 Rue Du Hammen 0", a date "1966-10-18", and a RIB number "486U03667". A central node with a question mark and ID "486U03667" is also connected to both main individuals. The interface includes a sidebar with navigation options like "Liste des alertes" and "Graphique du conflit", and a bottom status bar showing system information.

Exemple 1: nom et adresse proches, date de naissance et numéro de RIB identiques

IBM Relationship Resolution

Fichier Editer Afficher Aide

Boîte de réc... Conflits

Liste des alertes

Détail du conflit

Graphique du conflit

Recherche d... Entités Rapports UMF

ID alerte 42495 (Niveaux de séparation : 1)

Alerte sur : Entité 30267 - Entité 47793

Définir le statut Graphique Rapport

on	Règle de résolution	Score de résolution	Score de relation
C	Name & DOB	100	90

Chemin : Entité 30267 - Entité 47793

Détails de la correspondance

Entité : 30267			Entité : 47793		
Attributs	Code de source de données - Descripti...	ID externe	Attributs	Code de source de données - Descrip...	ID externe
Nom			Nom		
GLEN PETERSON STUART	C	22322	GLENN PERTERSON START		40102
Nom			Nom		
GLEN PETERSON STUART	C	22322	GLENN PERTERSON START		40102
DOB			DOB		
1966-10-18	C	22322	1966-10-18		40102
DOB			DOB		
1966-10-18	C	22322	1966-10-18		40102
BANK_ACCOUNT			BANK_ACCOUNT		
20 96U03667	C	22322	20041 J03667		40102

Entité 30267

Récapitulatif d'entité

Sources de données

Code de source de données - Descrip	ID externe	Date de création
C	22322	11/02/08 16:18

Attributs

Nom : GLEN PETERSON STUART

Adresse : 4 Rue Du t

DATE2: 12/21/06

DATE1: 12/28/06

GENDER: M

BANK_ACCOUNT: 2004

ID_CLIENT: 23455342

DOB: 1966-10-18

Entité 47793

Récapitulatif d'entité

Sources de données

Code de source de données - Descrip	ID externe	Date de création
C	40102	11/02/08 17:18

Attributs

Nom : GLENN PERTERSON START

Adresse : 4 Rue Du t

DATE2: 02/22/06

DATE1: 03/01/06

GENDER: M

BANK_ACCOUNT: 2004

ID_CLIENT: 26232017

DOB: 1966-10-18

Start 2 Windows Expl... notes.txt - Word... notes_day1.txt -... 3 Windows Co... IBM Relationsh... alertes_name_D... Console de confi... FR 78% 22:10

Agenda

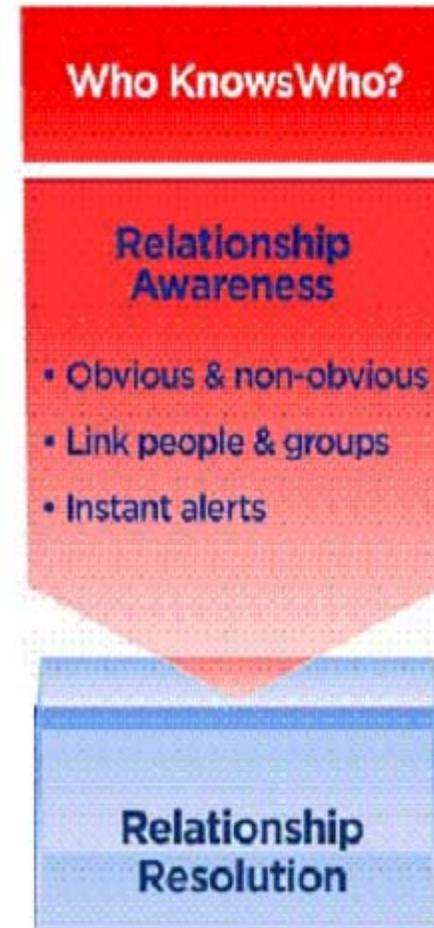
- EAS – GNR : une offre modulaire
- Entity Analytics Solution
 - Identity Resolution
 - Relationship Resolution
 - Anonymous resolution
- Global Name Recognition

Relationship Resolution

Relations évidentes ou masquées

Identification des relations entre personnes

Analyse proactive, pas de Data Mining nécessaire





Résolution de relations entre entités

Attributs d'identification

- Adresse
- N° d'immatriculation auto
- N°s de téléphone

Rôles

- Listes de personnes surveillées
- Réclamations/Cautions
- Listes noires
- etc...



Événements

- Personnes reliées par les lieux ou des événements
- Nombre importants de transactions dans une même journée par plusieurs personnes

Relations directes

- Contact en cas d'urgence
- Même adresse
- Même employeur

Relations indirectes

- Appels du même numéro de tél
- Utilisation de la même adresse postale
- Conducteur d'une voiture déclarée par quelqu'un d'autre

Résolution de relations entre entités

- Notions de rôles
- Définition de rôles incompatibles
- Génération d'alertes
- Audit



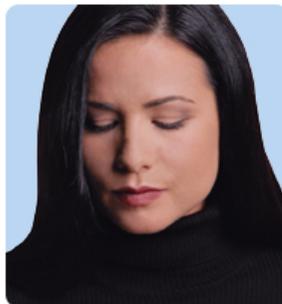
La résolution de relations – notion de “degré de séparation”

(transitivité: si $A = B = C$, alors $A = C$)



A: Renaud
Tél: 06 61 62 63 64

=



B: Annie
Tél: 06 61 62 63 64
Adr: 12 rue Haute

=



C: Thomas
Adr: 12 rue Haute



A: Renaud
Tél: 06 61 62 63 64

=



C: Thomas
Adr: 12 rue Haute

Renaud est en relation avec Thomas avec 2 degrés de séparation.

EAS gère jusqu'à 30 degrés de séparation!

Exemple 1: rapprochement de personnes partageant un même domicile

The screenshot displays the IBM Relationship Resolution interface. The main window, titled "Graphique de l'alerte", shows a network diagram of relationships. Two central nodes are highlighted in blue: "FREDERIC BARRANCO" (ID 18992) and "OLIVIER LOF" (ID 18991). They are connected by a thick blue line, indicating a relationship. Other nodes include a calendar icon (1964), a question mark (?), a date "13/08/97", a date "20/08/97", a date "12/08/97", a phone icon (5-6232), a house icon (Le Super Vallier, Mars, FRA, 13004), and another question mark (?). The left sidebar contains navigation options: "Boîte de réc...", "Conflits", "Liste des alertes", "Détail du conflit", "Graphique du conflit", "Recherche d...", "Entités", "Rapports", and "UMF". The bottom status bar shows "Aucun élément sélectionné". The Windows taskbar at the bottom includes the Start button, several open windows, and the system tray with the time 17:05.

Exemple 2: réseau partageant un numéro de téléphone commun

The screenshot displays the IBM Relationship Resolution interface. The main window, titled "Graphique d'entité - Entité: 41940", shows a network diagram where three individuals are connected to a shared phone number, 013-0. The individuals are:

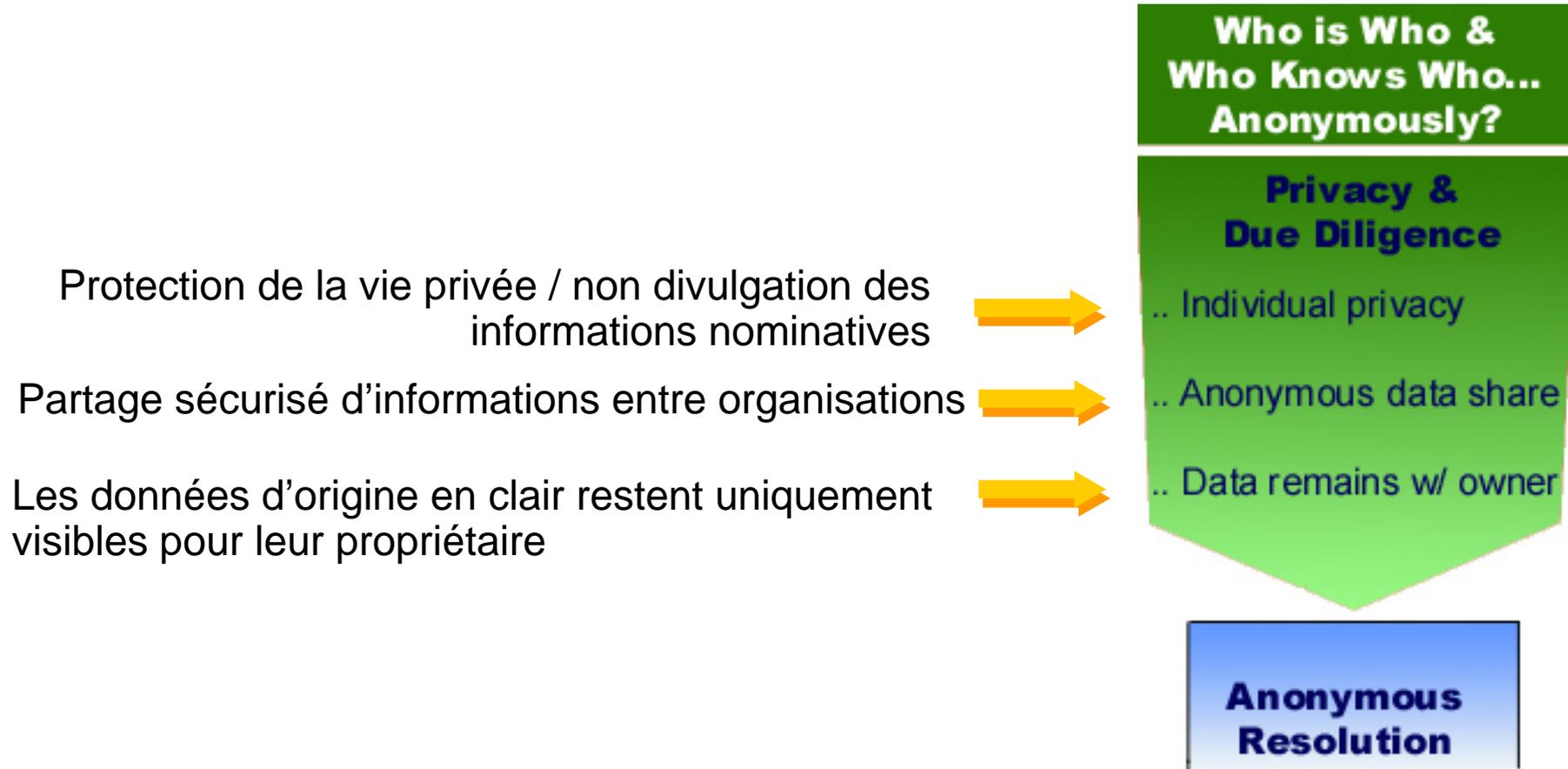
- JEAN PIERRE DINQUEL** (6049): Addressed at 91 B Av De La Maladrerie, Pois, FRA. He is connected to the shared number 013-0 and has two unknown relationships (marked with "?") with dates 17/10/06 and 30/10/06.
- NICOLE VERIHES** (41842): Addressed at 12 A Sentier Des Landes, C... F... She is connected to the shared number 013-0 and has two unknown relationships (marked with "?") with dates 15/07/06 and 01/08/06.
- ANITA NOEL** (41940): Addressed at Bat B 2EME Etage, Auben... She is connected to the shared number 013-0 and has two unknown relationships (marked with "?") with dates 21/11/06 and 16/11/06.

The shared phone number 013-0 is represented by a telephone icon. The interface includes a sidebar with navigation options such as "Ajouter une entité", "Récapitulatif d'entité", "Graphique de relation d'entité", and "Ajouter une divulgation". The bottom status bar shows "Aucun élément sélectionné". The Windows taskbar at the bottom indicates the system time as 14:49.

Agenda

- EAS – GNR : une offre modulaire
- Entity Analytics Solution
 - Identity Resolution
 - Relationship Resolution
 - Anonymous resolution
- Global Name Recognition

Anonymous Resolution

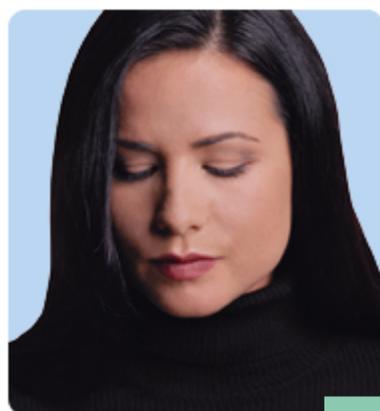




“Qui est qui et qui connaît qui... en respectant l'anonymat”

DB2 Anonymous Resolution

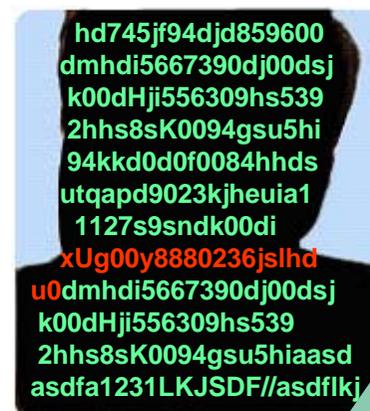
Permet à plusieurs intervenants de partager et de comparer des informations rendues anonymes.



Catherine Dupond
 1 rue de Bourgogne
 Chatillez 51700
 VIN# 585789543
Flying Blue: 5678965
 Tel: 03 49 05 60 55
 Passeport : 995027890



tr5y9hU000kdG563ks
 Hjd55603jd98hj4jfk9jt
 0089gehyud98kkdh00
 HydK8880h332jd78001
xug00y8880236jslhdu0
 0012g6743kd85hf06h7
 x084hdf75jc4539fhd89

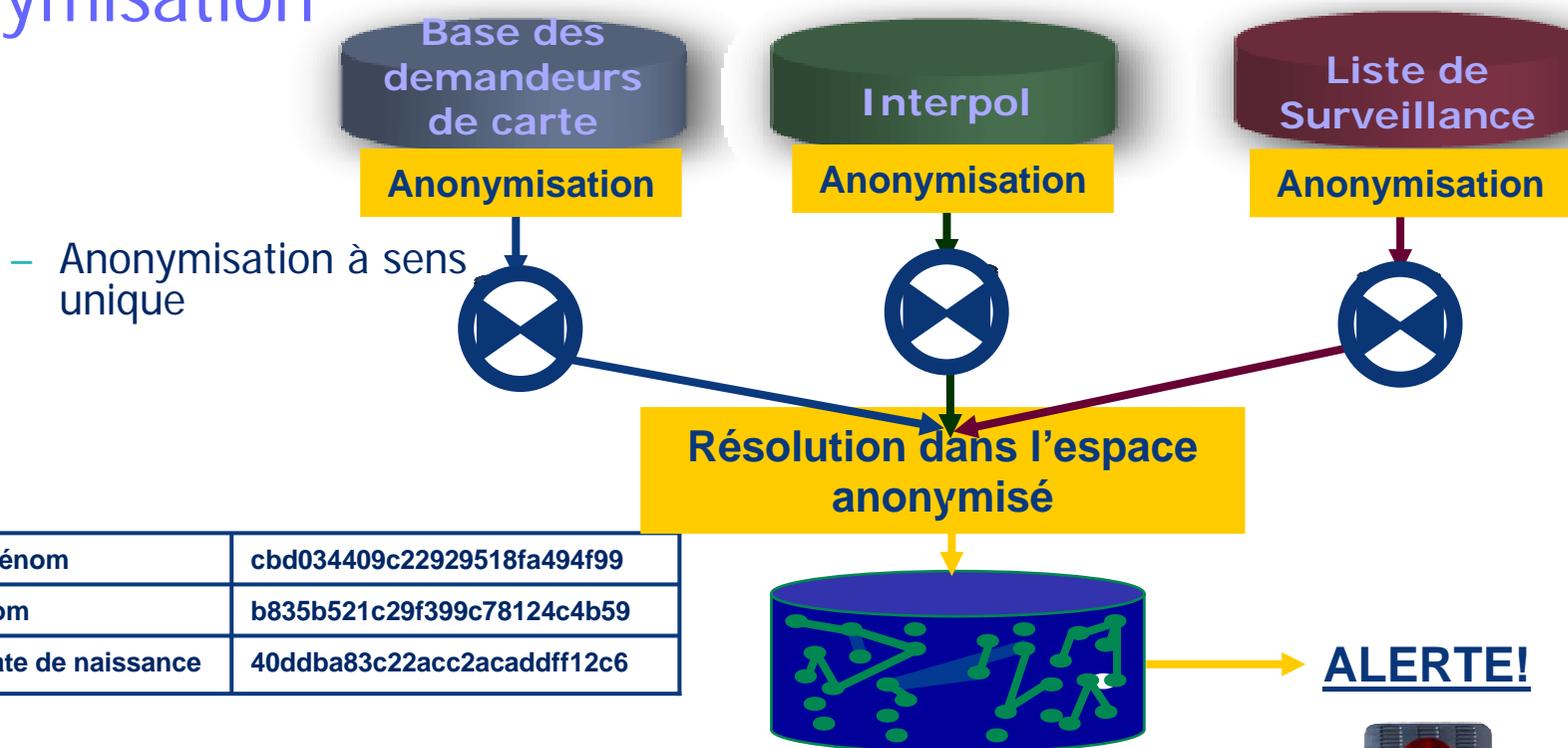


hd745jf94djd859600dm
 hdi5667390dj00dsjk00d
 Hji556309hs5392hhs8s
 K0094gsu5hi94kkd0d0f
 0084hhdsutqapd9023kj
 heuia11127s9sndk00di
xug00y8880236jslhdu0



Thomas Saint-Clair
 49 Rue Basse
 Louveciennes 78120
 Compte n° : 97836553122
 Compte n° : 00303450009
 Tel : 01 50 16 03 82
Flying Blue: 5678965

Fonctionnement et utilisation de l'anonymisation



- Pas de communication en clair de l'information
- Utilisation de l'information maîtrisée et sous le contrôle de son propriétaire
- Réduit les risques de violations des règles de protection de l'information



Architecture conçue pour des traitements en temps réel de grandes volumétries

- **Hautes performances**
 - Nombre illimité de sources de données
 - Résolution de milliers d'entités par seconde
 - Accès et résolution sur une base de données de centaines de millions d'enregistrements
- **Non intrusif**
 - Pas d'accès direct aux données sources
 - N'a besoin que des fichiers XML et des données mises à jour
 - Pas de compétences techniques additionnelles
- **Traitement temps réel**
 - Pas de mises à jour batch nécessaires
 - Auto-correctif
- **Capacité à traiter des volumes très importants de données**
 - Résolution sur de très grandes bases de données
 - La précision augmente avec la quantité de données
- **Temps de réponse faible et alertes en temps réel**
- **Opérationnel en 3 mois**
- **Pas de rupture de fonctionnement dus à des rafraîchissements ou rechargements**



Capacité

- >5,000 sources de données
- Gère >500 millions d'entités résolues
- Soit >3 milliards de lignes de base de données
- Débit (fil de l'eau) jusqu'à 2200 ERPS (Entités Résolues par seconde)

En conclusion : les points forts de la solution

- Utilisation de la totalité des attributs
- Conservation du contexte historique
- Détection de relations même très éloignées
- Performances élevées
- Génération d'alertes en temps réel

Agenda

- EAS – GNR : une offre modulaire
- Entity Analytics Solution
 - Identity Resolution
 - Relationship Resolution
 - Anonymous resolution
- Global Name Recognition



Reconnaissance des noms

Contrairement à d'autres éléments d'information, les noms revêtent différentes formes et particularités. Ce ne sont pas de simples chaînes de caractères, mais des objets flous susceptibles de beaucoup de variantes.

- ❑ Il n'existe pas de standard.
- ❑ Les noms peuvent contenir une variété d'informations optionnelles qui peut faire apparaître les noms de façon très différente :

Fautes de frappe,

- WEBER-VEBER

Bruit

- WEBER~W3BER

Titres

- Dr., Rev, Haj, Sri., Col.,...

Préfixes

- Fitz, O', De La, Abdul, ...

Surnoms

- Johnny, Betty, Alyosha, Paco, Drew

Phonétiques

- Leighton~Layton, Dena-Deane,
- Poudroux-Poudroux

Noms abrégés

- Fco, Ma, Mohd,...

Qualifieurs

- Jr., fils, neto, sobrinho, Ph.D.,

Séquences

- (Tung, Mao Tse)

Parties manquantes

- (Abdel ~ Abdel Rahman)

Variations

- (Mohammed, Imhemmed)





Exemple: Un nom identique dans différentes cultures Arabes

حاج محمد عثمان عبد الرقيب

1#. West Africa

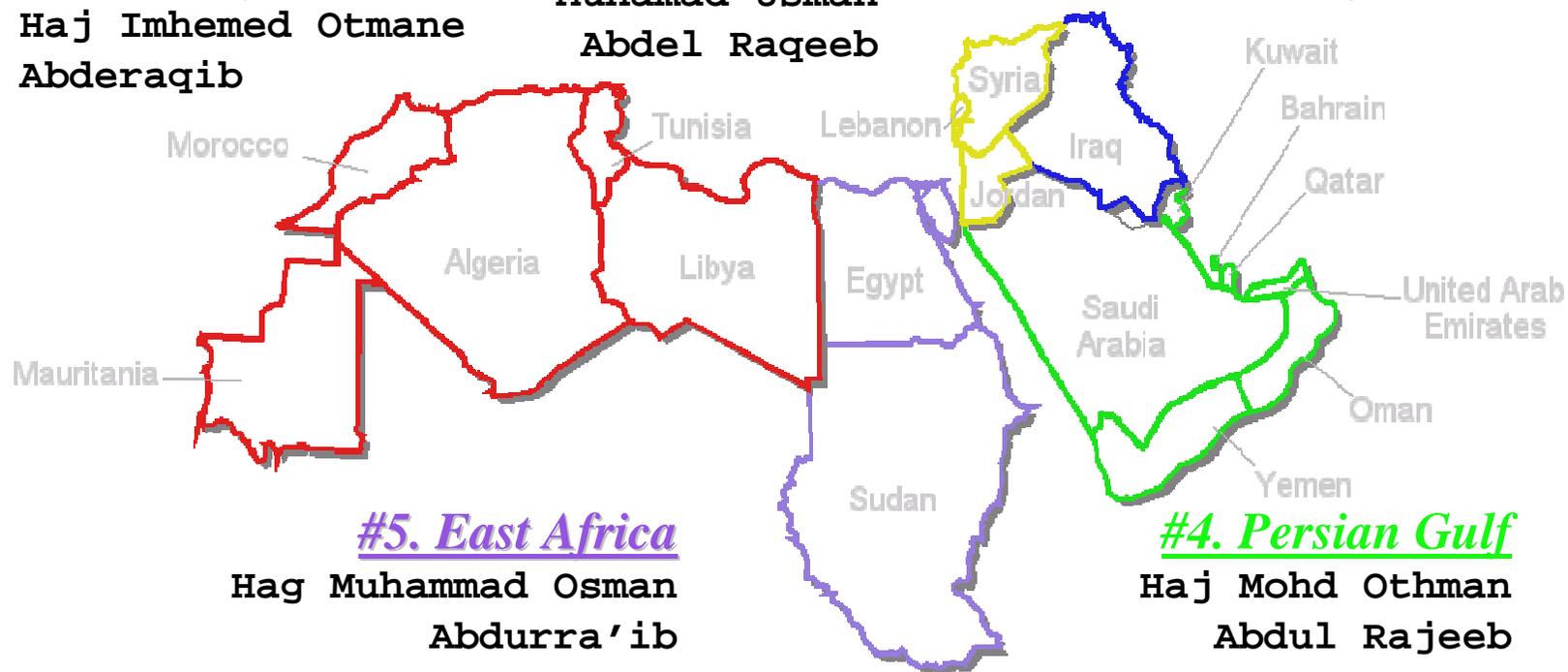
Haj Imhemed Otmane
Abderaqib

2#. Levantine

Muhamad Usman
Abdel Raqeeb

3#. Iraq

Hajj Mohamed Uthman
Abd Al Ragib



5#. East Africa

Hag Muhammad Osman
Abdurra'ib

4#. Persian Gulf

Haj Mohd Othman
Abdul Rajeeb

Exemple: le même nom Chinois en Asie du Sud-Est



張
丘
蘇

	Zhang Qiusu
	Chang Ch'iu-Su
	Sae Cheung, Yaw So
	Tiu, Koo Saw
	Zong, Kiu Su



Qu'est ce qui identifie un nom au sens GNR ?

- **Un *Nom*** qui identifie la famille ou origine de l'individu (simple ou composé)
- **Un *prénom*** qui identifie l'individu au sein de sa famille (initiale, prénom simple, composé...)
- **Un *surnom*** signalé sur une carte d'identité
- **Titres et qualificatifs** qui précèdent ou suivent le nom (Dr, Mme, Jr...)
- **Conjonction** (Mr et Mme Durand, Mr ou Mme Dupond)

Exemple: Dr. Edouard Louis Severson III

Comment GNR determine l'origine d'un nom ?

L'équipe GNR a étudié plus d' **1 milliard de noms** provenant de **205 pays différents**. Basé sur des similitudes linguistiques et culturelles, GNR s'appuie sur 20 groupes culturels différents pour classer les noms. La détermination de l'origine d'un nom permet au moteur de recherche d'utiliser les spécificités linguistiques du nom.

exemple : **Ziad Samir Jarrah...**

Name	Country	Rank
JARRAH (sn)	Lebanon	85
	United Kingdom	98
ZIAD (gn)	Lebanon	6
	Jordan	12
	Syria	12
	Saudi Arabia	31
	Tunisia	67
	Kuwait	73
	France	77
	Egypt	79
	Canada	86
	United Kingdom	87
SAMIR (gn)	Egypt	3
	Jordan	3
	Lebanon	4
	Syria	4
	Azerbaijan	8
	Bosnia and Herzegovina	9
	India	11
	Algeria	12
	Tunisia	13
	Saudi Arabia	22

Culture Details

Name	Farsi	Pakistani	Arabic	Afghani	Indian
Jarrah , Ziad Samir	100.000000	57.230000	28.997000	28.691800	28.586400

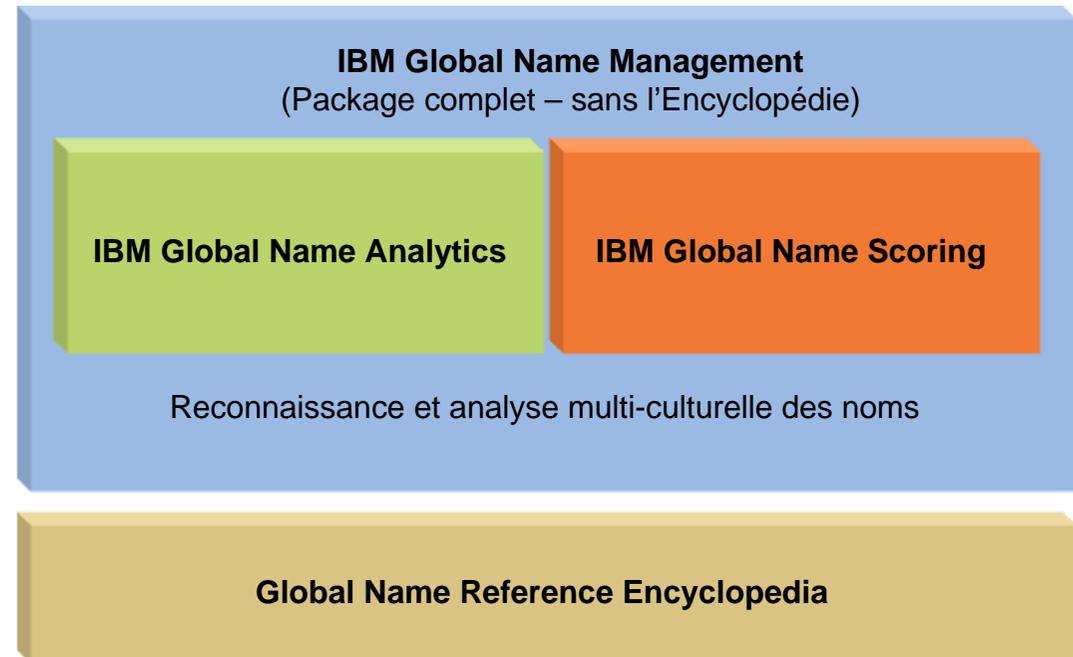
Surname	Farsi	Arabic	Hispanic	Pakistani
JARRAH	100.000000	0.385700	0.208300	0.203400

Given Name	Farsi	Pakistani	Arabic	Afghani	Russian
ZIAD	100.000000	100.000000	0.910900	0.339300	0.131400

Given Name	Arabic	Farsi	Indian	Pakistani	Afghani	Other
SAMIR	100.000000	100.000000	100.000000	100.000000	100.000000	0.000000

Global Name Recognition – Composants de l'offre

- **Global Name Management** – est une offre composée de:
 - **Global Name Analytics**
 - Name Works,
 - NameParser,
 - NameGenderizer,
 - NameClassifier,
 - Country of Association,
 - NameVariationGenerator
 - **Global Name Scoring**
 - Name Works,
 - NameHunter,
 - NameParser,
 - NameClassifier,
 - Country of Association,
 - NameSifter,
 - DateCompare.
- **Global Name Encyclopedia** – un outil de référence autonome fournissant des informations culturelles concernant l'usage des noms, leur sens et leurs modèles de variations orthographiques



GNR : détail des différents modules

- **GNR Name Works:**
 - APIs JAVA et SOAP haut niveau pouvant être intégrée à des applications ou appelée en Web Services
- **GNR Name Parser :**
 - Décomposition du nom global en nom de famille/prénom/suffixe/civilité..., suivant leur origine culturelle.
- **GNR Name Genderizer :**
 - Fournit le genre le plus probable d'un nom à partir de la fréquence d'attribution masculin/féminin dans des fichiers de noms dans des alphabets latin, chinois, coréen, hispanique et arabe.
- **GNR Name Classifier :**
 - Identifie et classe un nom par rapport à son origine culturelle; des noms traduits du Chinois, du Coréen, de l'Arabe peuvent présenter des variations orthographiques lorsqu'ils sont traduits en alphabet latin,
- **GNR Country of Association:**
 - Associe un nom à un ou plusieurs pays selon des règles statistiques
- **GNR Name Variation Generator :**
 - Calcule et fournit des variations sur un nom avec des scores de vraisemblance.
- **GNR Name Hunter:**
 - Fournit des services de recherche de noms avancés,
 - Application NameHunter Server pour la recherche dans de multiples listes de noms sur un serveur unique
 - Distributed Search Server pour la recherche dans de grandes listes de noms distribuées sur plusieurs serveurs
- **GNR Name Sifter :**
 - Sépare les noms de sociétés des noms de personnes
 - Permet le rapprochement des noms de sociétés
- **GNR Date Compare:**
 - DateCompare compare des valeurs de dates et retourne un score de similarité exprimé en pourcentage.

Classification culturelle des noms

- Identifie la culture d'un nom

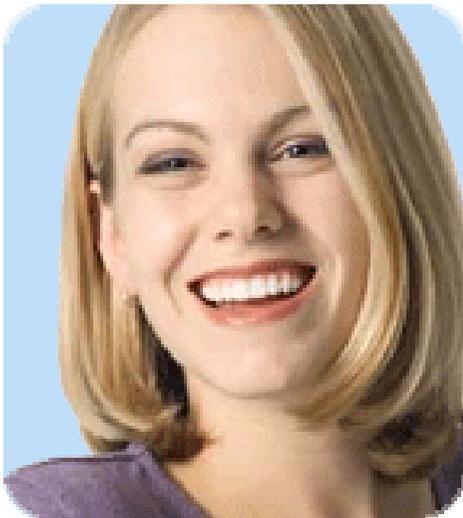


Entrée:

"MARIA LUZ RODRIGUEZ v DE LUNA"**Rang:****Classification****Nationalité****#1.****1074****Hispanique****#2.****534****Allemande****#3.****395****Anglaise**

Découverte du genre

- Identifie le genre d'un prénom



Entrée:

“Stéphanie”

Fréquence:

Probabilité

Pourcentage Féminin

99%

Pourcentage Masculin

0%

Scoring des noms

- Génère un score de comparaison des noms en fonction de la culture et du genre associés



Recherche:

“Bobby Creighton”

Retour:

Association à Creighton, Bobby

Noms:	Score:	ID:
Creighton,Robert	0.97	105875
Creighton,Robert & Margie	0.97	105876
Creighton,B J	0.955	105863
Mccreight,Robert E	0.69	487055

Modules de GNR: informations techniques

- Modules écrits en C++
- Appelés à partir de programmes d'application:
 - C ou C++ : appel direct via une bibliothèque .a
 - Compatible avec un environnement de développement Microsoft .NET
 - Name Works fournit une couche d'abstraction JAVA:
 - Appel à la librairie Java : via «connecteur» Java Native Interface
 - Dans n'importe quel Web Server J2EE 1.4:
 - ❖ Via un «connecteur» Web Services basé sur le protocole SOAP



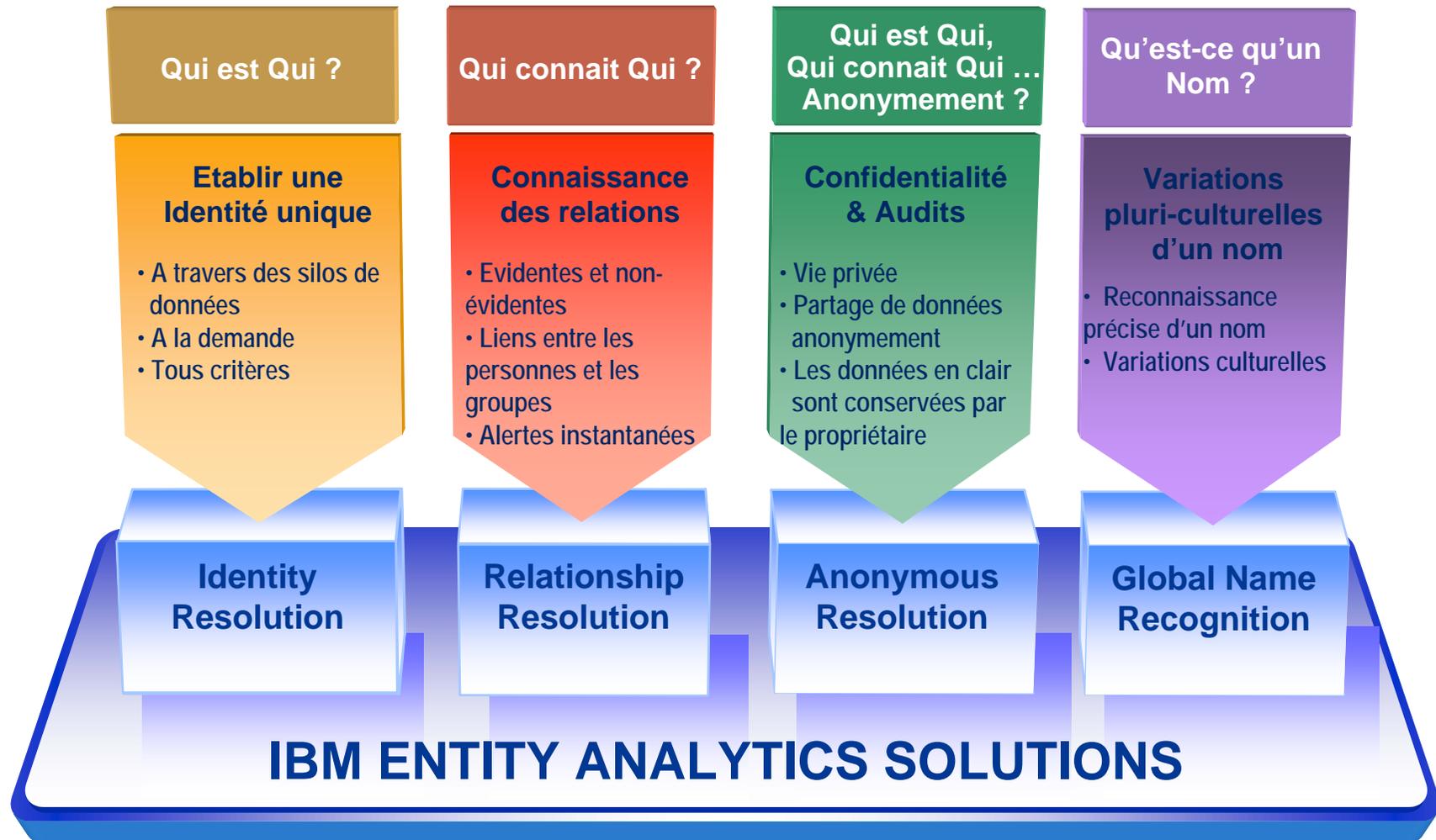
Les avantages de GNR

- **Amélioration profonde de la précision vis-à-vis des systèmes concurrents**
 - Réduit le nombre de résultats incorrects retournés (faux négatifs)
 - Prise en charge des informations culturelles (variations de noms)
 - Prise en charge de la phonétique (consonances, par ex: « é », « ai » et « ei »; « y » et « ill », etc...)
- **Plus de 200 pays couverts**
 - S'appuie sur des statistiques provenant d'une base de données contenant près d'un milliard de noms
 - Aboutissement de plus de 20 ans d'investigation et de recherche linguistique
- **Compatible avec tout type de configuration machine**
 - Algorithmes à haute performance (5 millions de noms /seconde)
- **Peut être utilisé sur les systèmes existants**
 - Mise en oeuvre rapide
 - Les interfaces de programmation et d'intégration permettent une intégration rapide dans le SI (JAVA, Web Services)

Plus de 200 pays étudiés pour GNR

-
- | | | | | |
|-------------------------|------------------------|----------------------|--------------------------|-----------------------------|
| 1 AFGHANISTAN | 42 CHINA | 83 GUINEA-BISSAU | 124 MARTINIQUE | 165 SAO TOME & PRINCIPE |
| 2 ALBANIA | 43 COLOMBIA | 84 GUYANA | 125 MAURITANIA | 166 SAUDI ARABIA |
| 3 ALGERIA | 44 CONGO (BRAZZAVILLE) | 85 HAITI | 126 MAURITIUS | 167 SCOTLAND |
| 4 ANDORRA | 45 COOK ISLANDS | 86 HONDURAS | 127 MEXICO | 168 SENEGAL |
| 5 ANGOLA | 46 COSTA RICA | 87 HONG KONG | 128 MICRONESIA, FED. ST. | 169 SEYCHELLES |
| 6 ANGUILLA | 47 CROATIA | 88 HUNGARY | 129 MOLDOVA | 170 SIERRA LEONE |
| 7 ANTIGUA & BARBUDA | 48 CYPRUS | 89 ICELAND | 130 MONACO | 171 SIKKIM (INDIA) |
| 8 ARGENTINA | 49 CZECH REPUBLIC | 90 INDIA | 131 MONGOLIA | 172 SINGAPORE |
| 9 ARMENIA | 50 DENMARK | 91 INDONESIA | 132 MONTSERRAT | 173 SLOVAKIA |
| 10 ARUBA | 51 DJIBOUTI | 92 IRAN | 133 MOROCCO | 174 SLOVENIA |
| 11 AUSTRALIA | 52 DOMINICA | 93 IRAQ | 134 MOZAMBIQUE | 175 SOLOMON ISLANDS |
| 12 AUSTRIA | 53 DOMINICAN REPUBLIC | 94 IRELAND | 135 NAMIBIA | 176 SOMALIA |
| 13 AZERBAIJAN | 54 EAST BERLIN | 95 ISRAEL | 136 NAURU | 177 SOUTH AFRICA, REP. |
| 14 BAHRAIN | 55 EQUADOR | 96 ITALY | 137 NEPAL | 178 SPAIN |
| 15 BANGLADESH | 56 EGYPT | 97 IVORY COAST | 138 NETHERLANDS | 179 SRI LANKA |
| 16 BARBADOS | 57 EL SALVADOR | 98 JAMAICA | 139 NETHERLANDS ANTILLES | 180 ST. CHRISTOPHER & NEVIS |
| 17 BELGIUM | 58 ENGLAND | 99 JAPAN | 140 NEW CALEDONIA | 181 ST. LUCIA |
| 18 BELIZE | 59 EQUATORIAL GUINEA | 100 JORDAN | 141 NEW ZEALAND | 182 ST. VINCENT & GRENADA |
| 19 BENIN | 60 ERITREA | 101 KAZAKHSTAN | 142 NICARAGUA | 183 SUDAN |
| 20 BERMUDA | 61 ESTONIA | 102 KENYA | 143 NIGER | 184 SURINAME |
| 21 BHUTAN | 62 ETHIOPIA | 103 KIRIBATI | 144 NIGERIA | 185 SWAZILAND |
| 22 BOLIVIA | 63 FALKLAND ISLANDS | 104 KOREA | 145 NORTHERN IRELAND | 186 SWEDEN |
| 23 BOSNIA | 64 FIJI ISLANDS | 105 KUWAIT | 146 NORWAY | 187 SWITZERLAND |
| 24 BOTSWANA | 65 FINLAND | 106 KYRGYZSTAN | 147 OMAN | 188 SYRIA |
| 25 BRAZIL | 66 FRANCE | 107 LAOS | 148 PAKISTAN | 189 TAIWAN |
| 26 BRITISH VIRGIN ISL. | 67 FRENCH GUIANA | 108 LATVIA | 149 PALAU, REP. OF | 190 TAJIKISTAN |
| 27 BRITISH WEST INDIES | 68 FRENCH POLYNESIA | 109 LEBANON | 150 PALESTINE | 191 TANZANIA |
| 28 BRUNEI | 69 GABON | 110 LESOTHO | 151 PALMYRA ATOLL | 192 THAILAND |
| 29 BULGARIA | 70 GAMBIA | 111 LIBERIA | 152 PANAMA | 193 THE BAHAMAS |
| 30 BURKINA FASO | 71 GEORGIA | 112 LIBYA | 153 PAPUA NEW GUINEA | 194 TOGO |
| 31 BURMA (MYANMAR) | 72 GERMANY | 113 LIECHTENSTEIN | 154 PARAGUAY | 195 TONGA |
| 32 BURUNDI | 73 GHANA | 114 LITHUANIA | 155 PERU | 196 TRINIDAD & TOBAGO |
| 33 BYELARUS | 74 GIBRALTAR | 115 LUXEMBOURG | 156 PHILIPPINES | 197 TUNISIA |
| 34 CAMBODIA | 75 GREAT BRITAIN | 116 MACAU | 157 POLAND | 198 TURKEY |
| 35 CAMEROON | 76 GREECE | 117 MACEDONIA | 158 PORTUGAL & AZORES | 199 TURKMENISTAN |
| 36 CANADA | 77 GREENLAND | 118 MADAGASCAR | 159 PUERTO RICO | 200 TURKS & CAICOS ISL. |
| 37 CAPE VERDE | 78 GRENADA | 119 MALAWI | 160 QATAR | 201 TUVALU |
| 38 CAYMAN ISLANDS | 79 GUADELOUPE | 120 MALAYSIA | 161 ROMANIA | 202 UGANDA |
| 39 CENTRAL AFRICAN REP. | 80 GUAM | 121 MALDIVE ISLANDS | 162 RUSSIA | 203 UKRAINE |
| 40 CHAD | 81 GUATEMALA | 122 MALI | 163 RWANDA | 204 UNITED ARAB EMIRATES |
| 41 CHILE | 82 GUINEA | 123 MARSHALL ISLANDS | 164 SAN MARINO | 205 UNITED KINGDOM |

Une plate-forme complète de résolution d'identités et de relations cachées



धन्यवाद

Hindi

多謝

Traditional Chinese

ขอบคุณ

Thai

Спасибо

Russian

Gracias

Spanish

شكراً

Arabic

Merci !

Obrigado

Brazilian Portuguese

Grazie

Italian

多谢

Simplified Chinese

Danke

German

Thank You

English

நன்றி

Tamil

ありがとうございました

Japanese

감사합니다

Korean