# 7 Things
## to Look for in a Name Matching Solution

**The must-have features you need to
achieve quick, effective name matching**

# What Makes an Effective Name Matching Solution?

Name matching presents complex challenges for actions ranging from pulling up a customer record, to verifying identities and watchlist screening for border security, to financial compliance. Unlike other forms of data, names are highly variable – spelling, misordered components, and even nicknames can result in a missed or false match. Traditional methods for matching names, like building massive lists of name variations, are simply too slow and can't scale to today's fast-paced environment.

Modern, AI-driven fuzzy name matching solves many of these issues, and it does so with accuracy, speed, and efficiency. But what should you look for when selecting a name matching solution to meet your needs? This ebook discusses the most important criteria to consider.

**Here are the most important criteria to use when selecting a name matching solution to meet your needs.**

## A primer on fuzzy name matching

Do you need to brush up on the basics of name matching?

[Download The Complete Guide to Name Matching](#) to learn more about the challenges it presents, and the ways that technology has advanced to overcome them in different domains. You'll also learn how AI-powered hybrid two-pass matching combines multiple technologies to maximize accuracy and speed across a broad range of name variations.

# How to Evaluate Name Matching Solutions

Businesses and agencies using name matching must weigh seven considerations when it comes to choosing a solution, including speed, accuracy, and cost. Each of these affects the other in direct and indirect ways. For example, speed and accuracy typically have an inverse relationship. Greater accuracy requires more processing time, and vice versa. Accuracy is also not black and white. There are different types of accuracy depending on your specific business priorities. Do you need to find every match or do you need to minimize false matches?  Your answers will help determine the weight of your evaluation criteria. Whatever your business needs, here's what you should think about:

**1**
**Accuracy**

**2**
**Speed and performance**

**3**
**Multilingual and multicultural support**

**4**
**Explainable AI**

**5**
**Easy tuning and customization**

**6**
**Matching with other identity attributes**

**7**
**Flexibility**
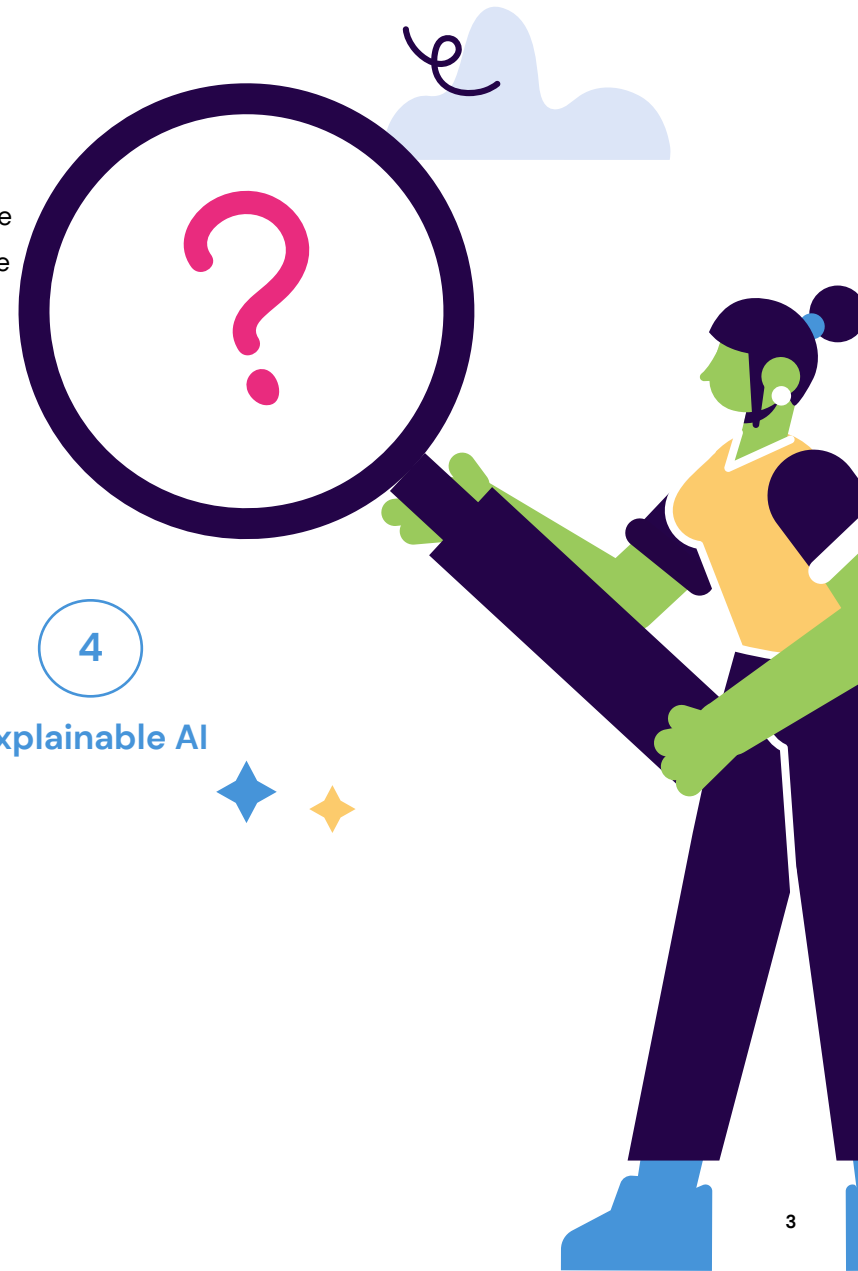
# ① Accuracy

**The top three requirements are accuracy, accuracy, accuracy.** It's tempting to ask "What is your tool's accuracy?" but that question is meaningless. The question you should be asking is "How accurately can this tool perform on my data?" High accuracy on arbitrary data isn't a good test. Test data doesn't need hundreds of examples, but it does need to contain name variations your current solution handles well, doesn't handle well and all other types of name variations in your data.

Furthermore, accuracy is not an absolute, but rather a balance between precision and recall that requires adjustment based on use case. Best-in-class name matching solutions allow you to adjust the mix of precision and recall for your use case, and tune the tool to your data's needs to maximize overall accuracy (see point #5 below).

## Understanding accuracy, precision, and recall

Accuracy is calculated as the harmonic mean of precision and recall. It's called the F-score which ranges between 0 (low) to 1 (high), and can also be expressed as a percentage.

**Precision** measures the percentage of matches that are correct. That is, high precision means very few false positives (incorrect matches).

**Recall** measures the percentage of all possible correct matches that the system found. That is, high recall means very few false negatives (missed matches).

# The relationship between precision and recall

Generally, precision and recall have an inverse relationship: the higher one is, the lower the other. Accuracy can be driven by higher recall or higher precision.

**FOR EXAMPLE**

When accuracy (F-score) is 75%, it could be the case that:

precision = 90% and recall = 64%
                    *or*
precision = 70% and recall = 81%.

Use cases such as finding patient records, directory searches, watchlist screening, or intelligence analysis require high precision so users aren't overwhelmed by looking through many false positives. Use cases in which the consequences of missing a match are most dire, such as no-fly lists, require high recall, so that missing a match is less likely. Maintaining precision while maximizing recall lets border security process most travelers quickly, while diverting higher-risk travelers for further review.

## WHAT TO ASK YOURSELF

Does your use case need higher precision to reduce false positives (non-matches labeled as a match) or higher recall to not have false negatives (missed matches)? And does this solution let me choose?

# 2  Speed and Performance

**In certain environments, time is of the essence.** Border security agents only have a few minutes with each passenger, so a slow response to a search will quickly cause a logjam. In an infrastructure-constrained environment (e.g., power outage at an airport), you might have to check passengers against watchlists with millions of names on a laptop instead of a powerful server.

Few organizations can afford the time lost to slow, inefficient name matching. Take for example, onboarding new customers to a cryptocurrency buying platform. If they get delayed by a false positive hit on a watchlist, they may go to a competitor to complete their time-sensitive transaction. Increased speed also means being able to keep up with the exponentially increasing number of banking transactions that need screening against watchlists daily.

## WHAT TO ASK YOURSELF

Can the solution give me the throughput I need?
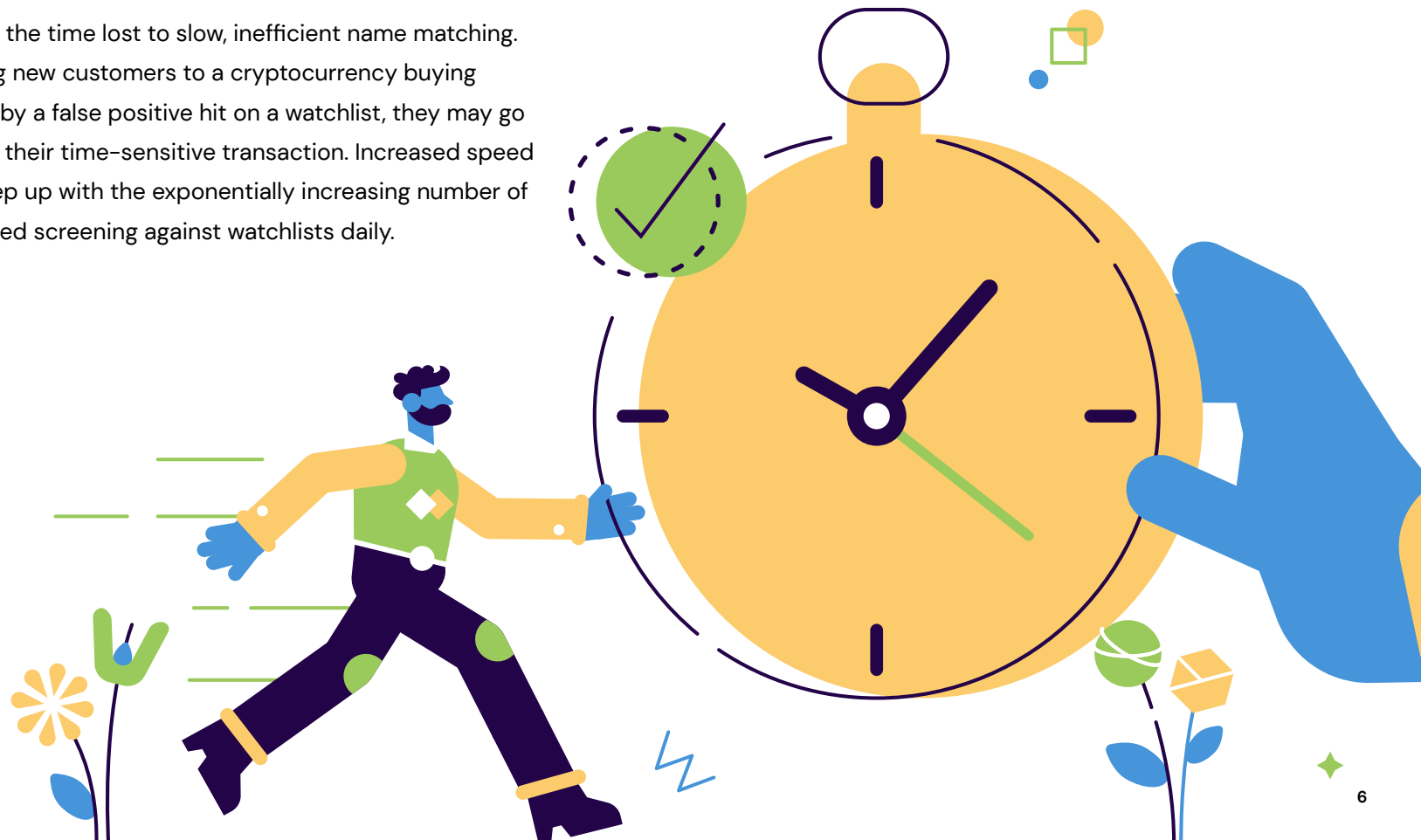
# (3) Multilingual and Multicultural Support

**Your names may all be written in English, but especially in the U.S., names originate in countries and languages from all over the world.**

For example, Arabic names transliterated into English are spelled myriad ways (e.g., Gadaffy, Khadaffi, Qaddafi). The knowledge to map transliterations back to the original Arabic name increases matching accuracy. Doing business in a foreign country, you may also have to screen against watchlists in the native language. A Japanese watchlist might contain Arabic-origin names transliterated into Japanese.

Additionally, understanding which names are male or female requires cultural knowledge. "Jean" in English is primarily a female given name, but in French it's primarily masculine, such as in Jean-Luc. Thus, if you are matching a French "Jean" to an English "Jean," there may or may not be a gender mismatch.
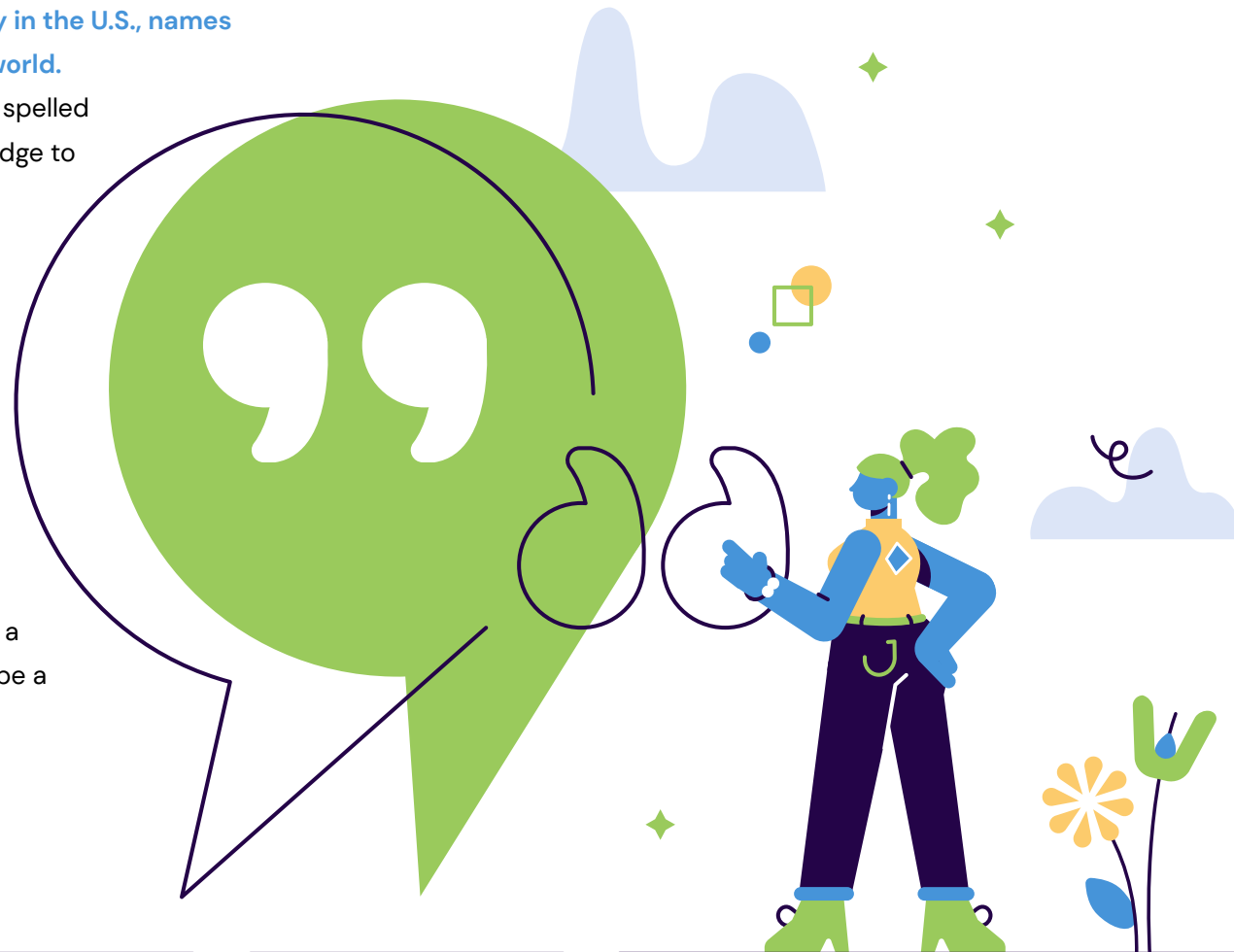
# Name characteristics differ by language

Approaching multilingual names with a linguistic and cultural understanding of the native language produces the best results. Here are just a few examples:

**Japanese** names use the same Kanji characters as in prose, but as a name, the pronunciation is often completely different. The character for east 東 pronounced "toh" or "higashi" is pronounced "Azuma" as a family name.

**Chinese** speakers write foreign names in Chinese ideographic characters (Hanzi), but they may vary widely depending on the region. "Portman" is written:

寶雯 (baowen) in Hong Kong

波特曼 (boteman) in China

波曼 (boman) in Taiwan

**In Spanish–speaking countries,** many people have two surnames: maternal and paternal.[1] In databases, one of the surnames may be omitted or input into the middle name field by accident.

**Thai** children are commonly given a nickname (not derived from their given name) in addition to a given name. In databases, often a nickname may be mistakenly labeled as the given name, and the given name labeled as a middle name.

## WHAT TO ASK YOURSELF

Does the solution handle variations from transliteration? What about language– and culture–specific naming conventions? If applicable, does it match names written in different language scripts?

[1] Dos Apellidos: When Families Have Two Surnames, Nicolas Cabrera, November 2020.

# 4 Explainable AI

**Another important consideration when evaluating a solution is its approach to artificial intelligence (AI).** AI is often a "black box" to anyone except its developers. It produces results without obvious reasoning, and users must simply trust that they're accurate. That might be fine for web searches, but in many name matching cases, it's important for the user to know how the AI works for operational reasons.

Financial compliance regulators or government entities typically require a clear explanation for a name match. For example, why was someone found ineligible for healthcare subsidies? That means you need a solution that offers transparency and explainability around its use of AI. It should have the ability to reveal — in granular detail — how it arrived at a match and why it calculated a certain score, both for business users and data scientists.

As regulatory requirements continue to shift or data evolves, you may need to adjust match behavior and to do so, you have to know how your solution's AI is determining whether two names match. Otherwise, it's impossible to customize matching behavior to work with new policy, regulations, or data idiosyncrasies.
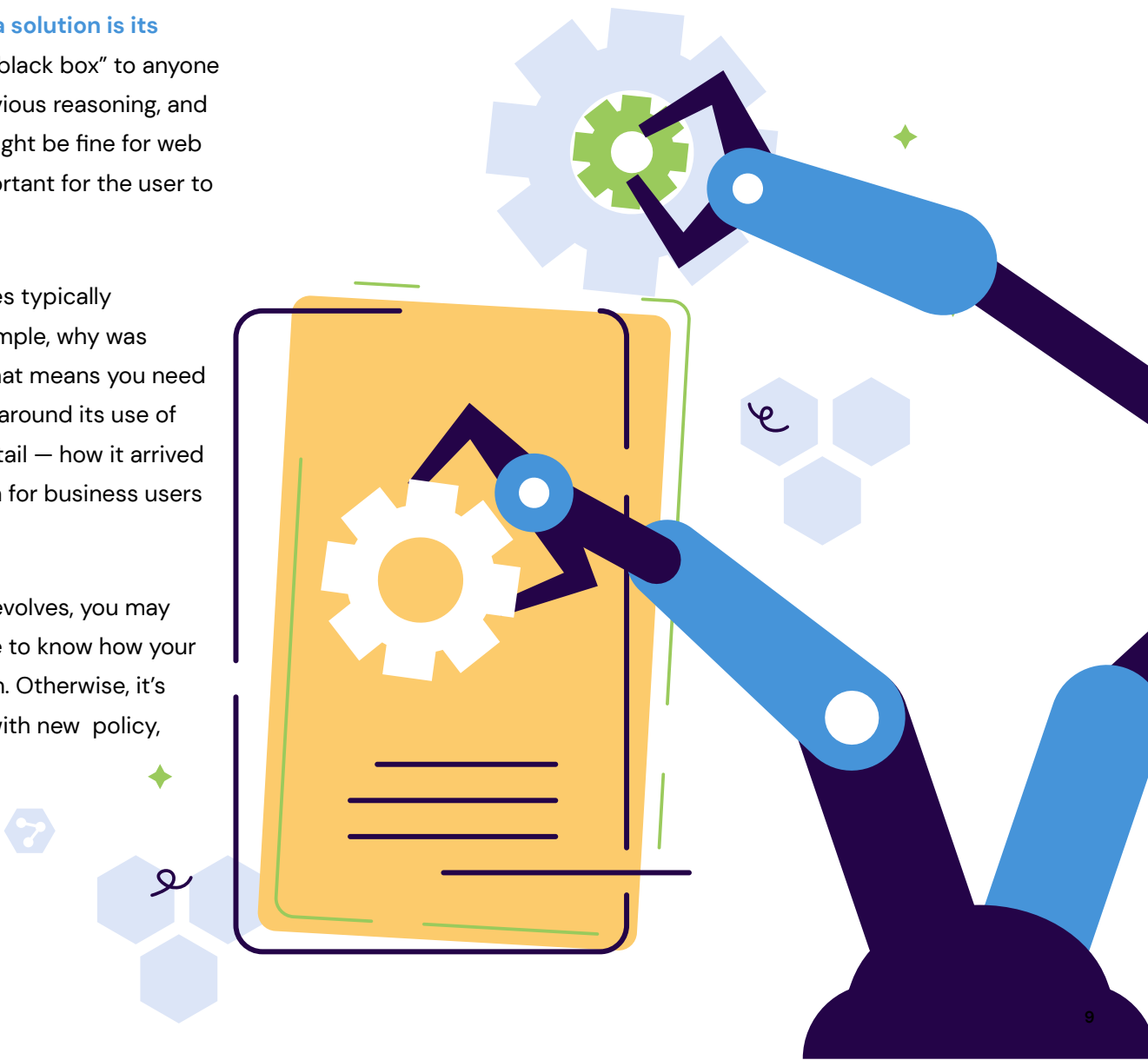
# ⑤ Easy Tuning and Customization

**There's a lot that a best-in-class name matching solution can do out of the box, but that doesn't mean it can't be improved upon.** You should be able to adapt your solution's matching behavior to your data set and use case, such as reducing recall to increase processing speed. Look for a solution with a user-friendly interface for non-technical business users to be able to see the impact of tuning changes on match behavior in real time.

Users should be able to weight different identity fields depending on their reliability, and manually set equivalent terms or names specific to your use case, for example:

Car ↔ auto (equivalent terms)

Alexander ↔ Sandy (nicknames)

Stefan ↔ Steve ↔ Etienne (cognates)

# Customize criteria for your use case

Watchlist screening providers often need the ability to set different match configuration profiles for each organizational client and to adjust match criteria according to regulatory changes. This occurred recently when the U.S. Office of Foreign Assets Control (OFAC) changed the match requirement for entities on the non-sanctioned Chinese Military Industrial Complex list from "exactly matches" to "exactly or closely matches."[2]

Suppose your database includes many names of Spanish origin. You may wish to lower the penalty on a missing name component if one database is often missing one of a two-word surname.

## WHAT TO ASK YOURSELF

Do I get a match score that enables me to set a threshold score for what is a match? Can a business user adjust match behavior and scoring to fit my data and use case? What types of variations do I see in my data and can I configure this solution accordingly?
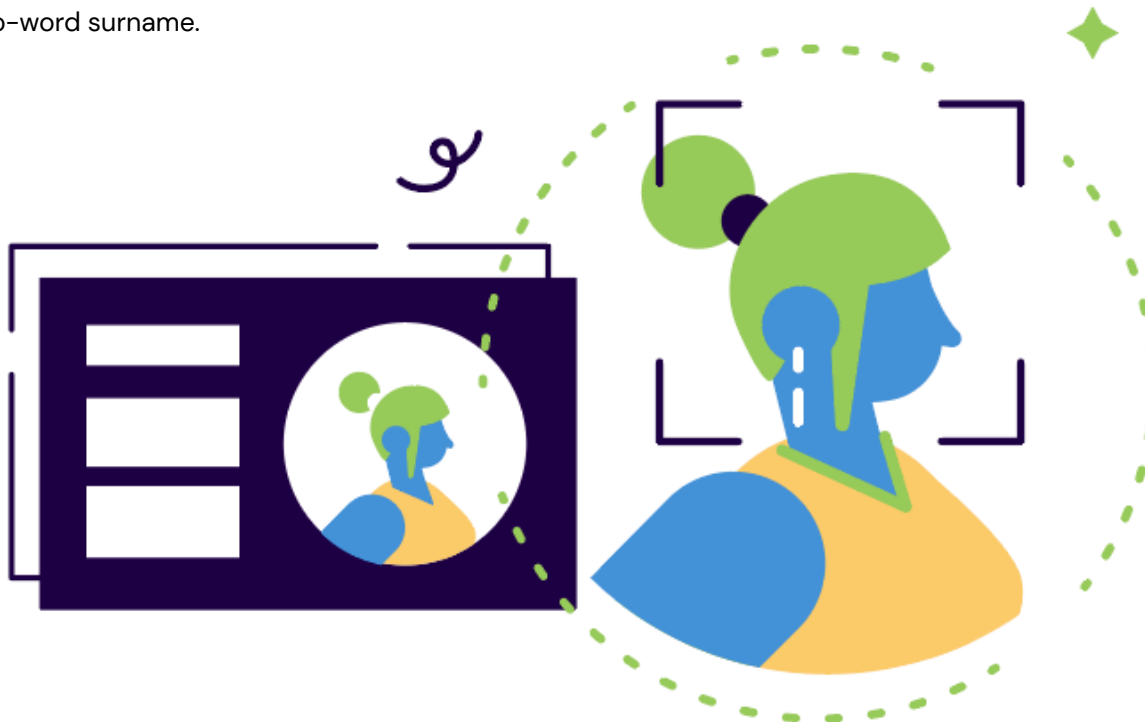
[2] "White House Issues Amended Executive Order on Chinese Military-Industrial Securities," June 4, 2021, Steptoe International Compliance blog.

# (6) Matching with Other Identity Attributes

## ADDRESSES

Capabilities to look for include parsing unfielded addresses and applying distance calculations to fields using postal code or street number. It should understand address-specific abbreviations, as well as the meaning of postal codes. In addition, your solution should apply its name matching algorithms to fields such as "street name," "building name," "city," "province/state," and "country."

## DATES

Like names and addresses, dates have certain idiosyncrasies that your solution should be able to handle. It should be able to compare partial dates and differently-ordered date components (DDMMYY versus MMDDYY). And it should be able to calculate the difference in days, months, or years between two dates, as well as understand chronological proximity. Consider that June 30 is closer to July 1 than July 30, despite the numerical similarity, or that Dec 31, 1969 is chronologically close to Jan 1, 1970.
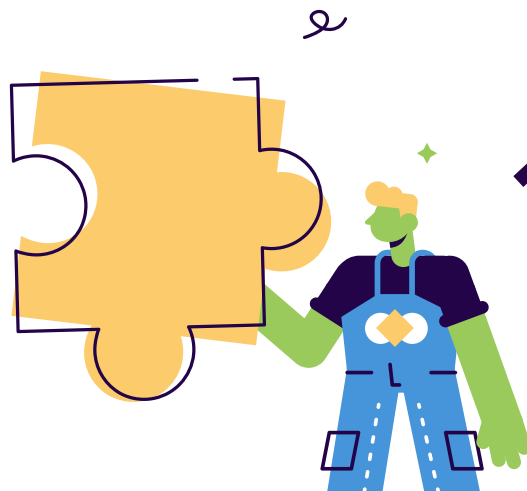
## OTHER IDENTITY DATA

It's not just addresses and dates, either. You may want to configure your solution to analyze tattoos, car VINs, known associates, and whatever else you need. When those needs pop up, you don't want to find out your solution isn't up to the challenge. You should know from the start that it has the ability to handle anything.

### WHAT TO ASK YOURSELF

If I have other identity fields, does this solution fuzzy match those field types? Can they be incorporated into the final match score?

# ⑦ Flexibility

**To grow and scale your business without problems, you'll need a flexible name matching solution that can meet your needs now and in the future.** The right fit will have a big impact on your results, but not your deployment, operating complexity, or budget.

Everyone's worst nightmare is a costly, complicated rip-and-replace project — especially since you might be happy with most aspects of your current systems, just not name matching. Some deployments can drag on for months and require countless hours of professional services.

It doesn't have to be that hard. You should be able to find a name matching solution that fits right in with what you already have with minimal effort — like writing a few lines of code to call an API, instead of costly hardware tie-ins. Integration and deployment should be possible in days or hours, not months.

## Pro tip for integration

Look for seamless integration paths to your system. Especially if you're running popular solutions such as Elasticsearch, Salesforce, or Solr, there should be plugins and extensions that let you take advantage of advanced fuzzy name matching without leaving your other solution.
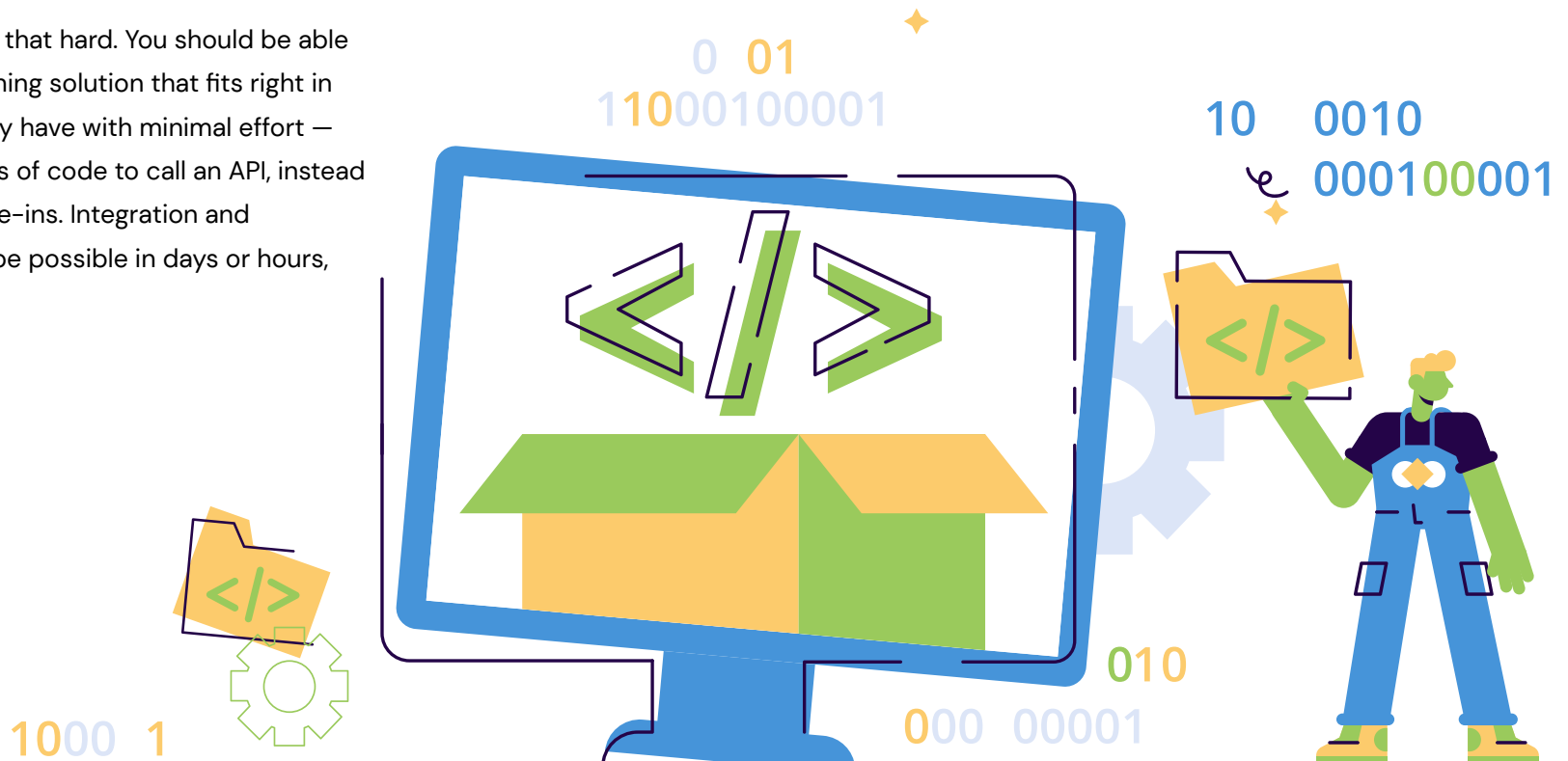
# Flexible licensing with no hidden costs

It's not just deployment that can be a frustrating, inflexible process. Many solutions still rely on rigid, antiquated licensing models which only scale up in large, expensive tiers that tie capacity to hardware purchases. Needing 5% more capacity might entail buying 50% more before you need it because of the hardware investment.

That's not right. Instead, you should be looking for a flexible licensing model that's based on what you actually use with no hidden costs. That way, you have full control over your infrastructure investment. You'll be able to understand and predict how costs will change as you grow, helping you scale your business effectively.
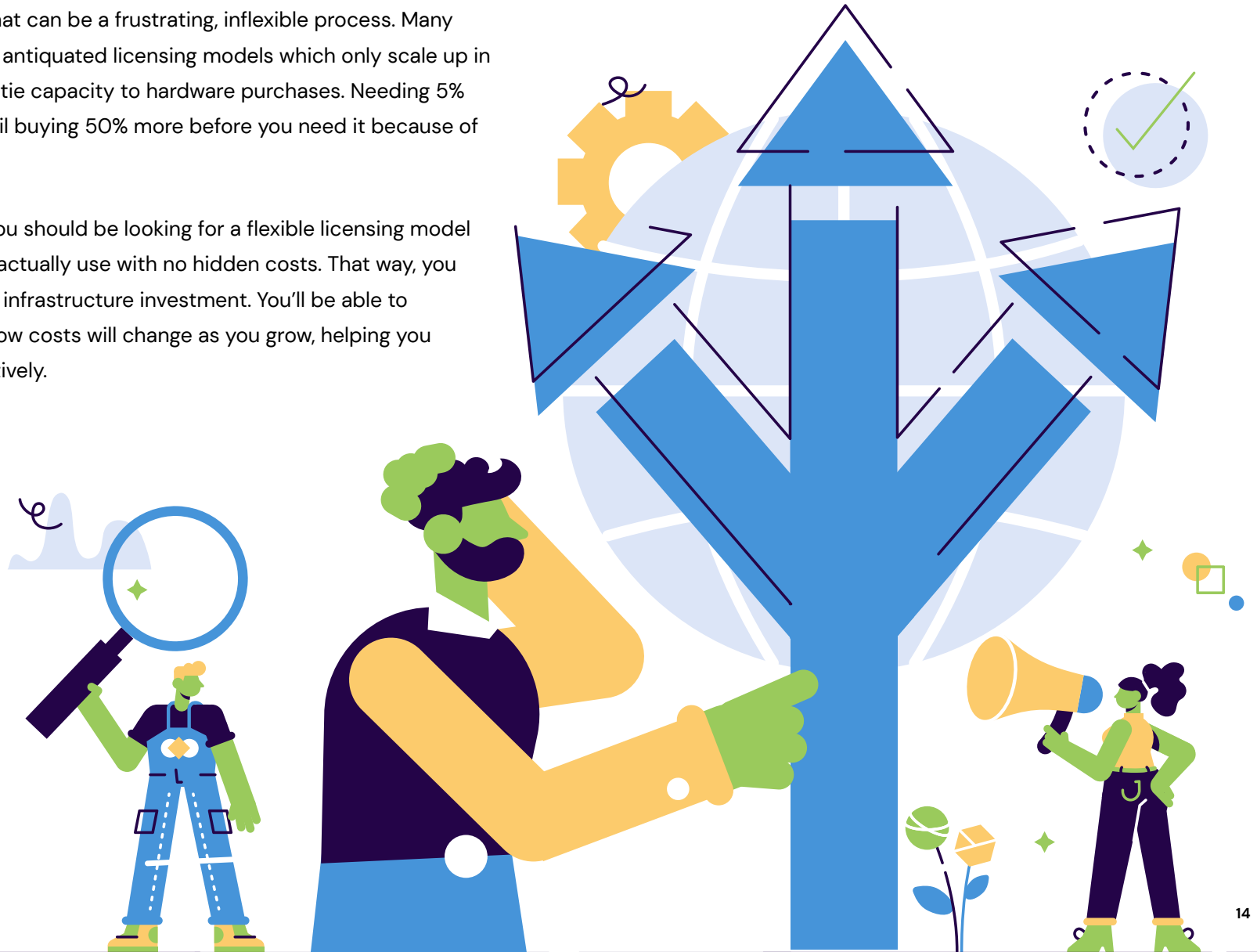
# And One More Critical Consideration... the Vendor

Features and capabilities are crucial criteria when selecting a name matching solution, but there's one more that is no less critical: the vendor itself. Name matching is highly complex, and the technology is evolving. You deserve to work with a solution vendor who has earned your trust. Look for thought leadership and industry credibility.

How long has the vendor been in business? Is name matching the company's core competency or just a lower-tier product in its lineup? Is it actively developing the product with regular updates and substantial improvements, or is it gathering dust? Given the fast-changing nature of regulations and data, does the vendor listen to and respond to customer needs? And finally, how responsive is its customer support? Does it provide access to developers, or just customer-support gatekeepers? Does it add features to the product roadmap based on customer feedback?

**These are all questions that you need answered — after all, even if a company has what looks like a good solution now, you need assurance that the solution will evolve and the vendor will be there to support you going forward.**

# In Conclusion

When selecting a name matching solution, the best approach is "quiet disruption." That means choosing a solution that vastly improves your matching process (and thus ROI) without disrupting your operations. Best-in-class matching solutions don't replace your current applications. Good solutions make applications much more effective by integrating with just a few lines of code. Your best-case scenario is to find a solution that doesn't require rip-and-replace — one that can disrupt results in a good way without interrupting your work.

And finally, the best solution will have a high ROI that reduces total cost of ownership by providing increased effectiveness and productivity, lowering operational costs, and letting your team focus its time and resources on your core business. That requires a powerful, purpose-built solution that is easy to integrate, adapts to changing needs, and is easy to manage.

# Rosette for Identity Intelligence

Rosette® by Babel Street, a text analytics platform by Babel Street, leverages machine learning to perform fuzzy name matching. Our approach reduces false positives and false negatives and has the flexibility to be uniquely tuned to perform optimally on each use case and data set. Rosette isn't limited to Latin script languages: it currently supports multiple complex languages, and resolves numerous name variation issues. In addition to names, Rosette can handle dates, addresses, and other identity attributes. Rosette is trusted by leaders in the financial and government sectors.

## Babel Street. Unlock the most insights that matter.

Babel Street is the trusted technology partner for the world's most advanced identity intelligence and risk operations. The Babel Street Insights platform delivers advanced AI and data analytics solutions to close the Risk-Confidence Gap.

Babel Street provides unmatched, analysis-ready data regardless of language, proactive risk identification, 360-degree insights, high-speed automation, and seamless integration into existing systems. We empower government and commercial organizations to transform high-stakes identity and risk operations into a strategic advantage.

Learn more at babelstreet.com

**BABEL**STREET