



## Seven Smart Data Rules - to keep your data clean and usable

Keep your database tidy and usable long-term

29 January 2019



# How does clean data help you?



Better productivity



Enhanced accuracy and trust in your data



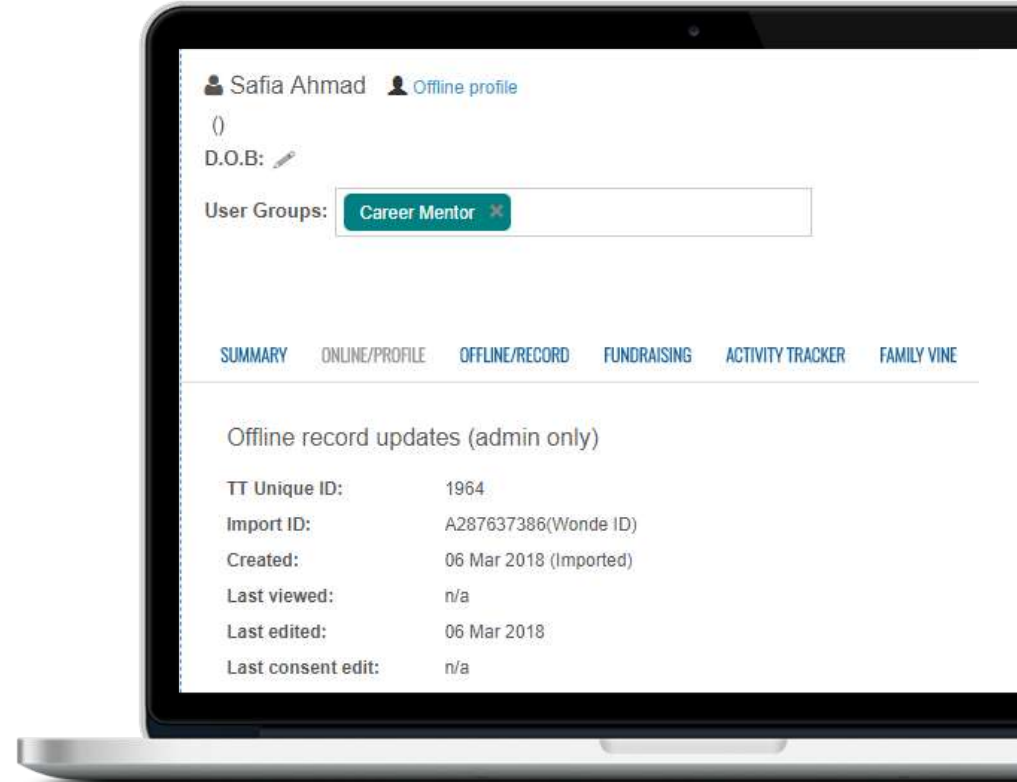
Helps you make more informed decisions



Improves the efficiency of your internal processes

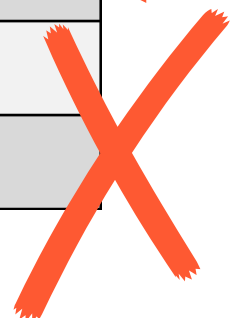
# Rule 1: Always use unique IDs

- ★ **Unique number/ text string**, clearly identifies a record in your database.
- ★ **Record**, is made up of parts which might not be unique to the individual, The identifier makes each person unique
- ★ These can be **serial** (assigned incrementally/sequentially) or **random**



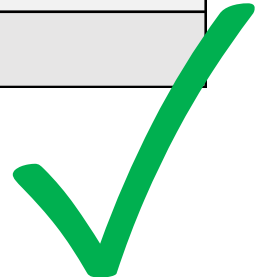
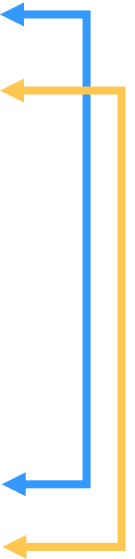
First Name	Last Name	Start year when at school	End year when at school
John	Smith	1987	1991
John	Smith	1987	1991
Lucy	Edwards	1988	1991
Mary	Stephens	1988	1991

Connection to institution	House	Class of	Willing to give career advice
Former Student	Bixby	1993	Yes
Former Student	Yacht	1991	No
Former Student	Hatton	1991	Yes
Former Student	Hatton	1991	Yes



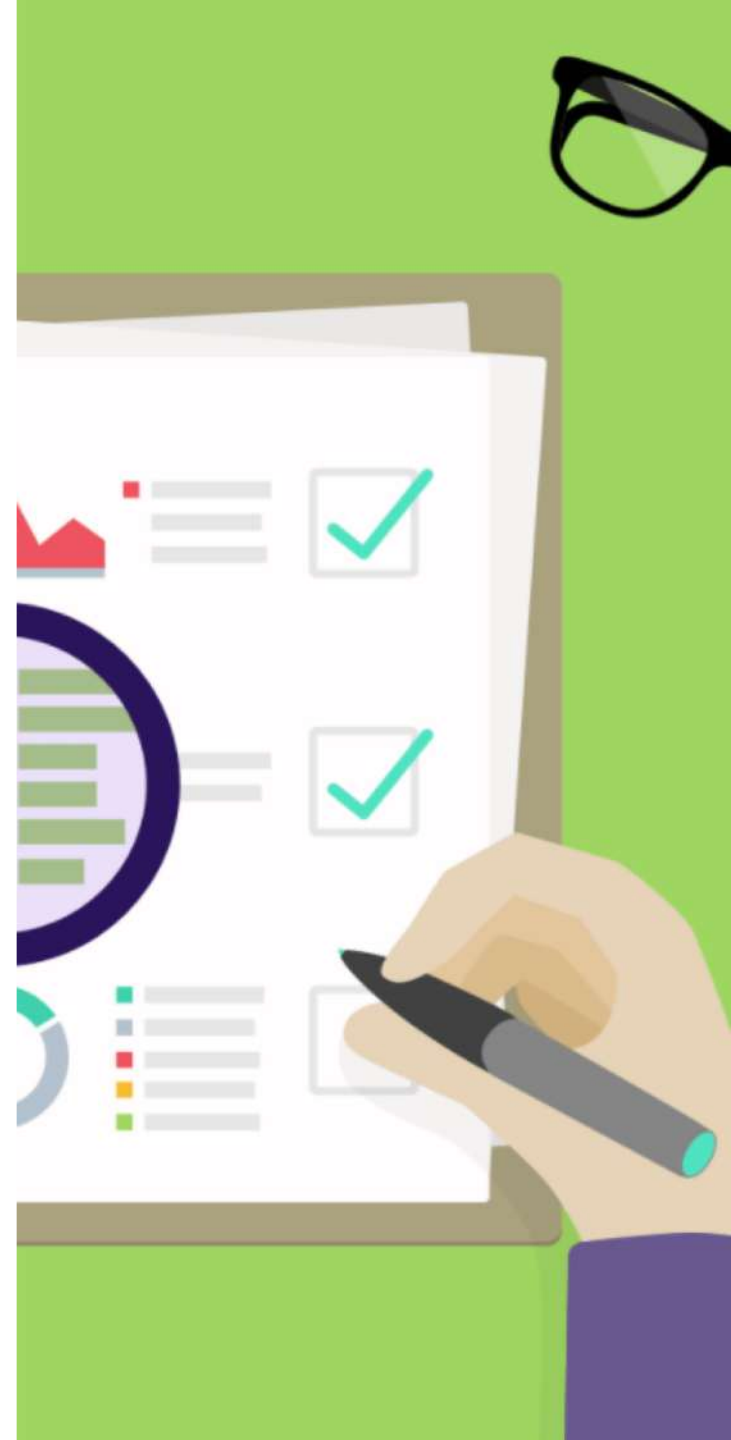
Unique ID	First Name	Last Name	Start year when at school	End year when at school
23748923658239	John	Smith	1987	1991
72839472839383	John	Smith	1987	1991
23423123892339	Lucy	Edwards	1988	1991
72833492738493	Mary	Stephens	1988	1991

Unique ID	Connection to institution	House	Class of	Willing to give career advice
23748923658239	Former Student	Bixby	1993	Yes
72839472839383	Former Student	Yacht	1991	No
23423123892339	Former Student	Hatton	1991	Yes
72833492738493	Former Student	Hatton	1991	Yes



# Rule 2: Validation is your friend

- ★ **Data validation** essential in making sure data you capture and maintain is clean and logical.
- ★ **Validation** is an automatic check to ensure that the data entered is sensible, reasonable and serves a good purpose
- ★ Use **fixed lists** to capture data (multiple choice or single answer) ensures your fields are consistent & usable.
- ★ **Limiting data** to a set of values/options allows you to record your data in a sensible and logical manner, allowing you to easily search, interpret, understand and report on your data.



# Fixed lists & custom questions

Subjects taught

-- Select an option --

Current Town/City

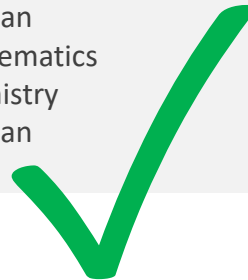
Current Country

Current Employer

- English Literature
- Mathematics
- Chemistry
- Biology
- Physics
- English Language
- French
- German

Mathematics  
English Literature  
Chemistry  
Biology  
Physics  
English Language  
Physics  
French  
Mathematics  
French  
German

Physical Education  
Biology  
Biology  
French  
English Literature  
English Literature  
French  
German  
Mathematics  
Chemistry  
German



# Free text

Subjects taught



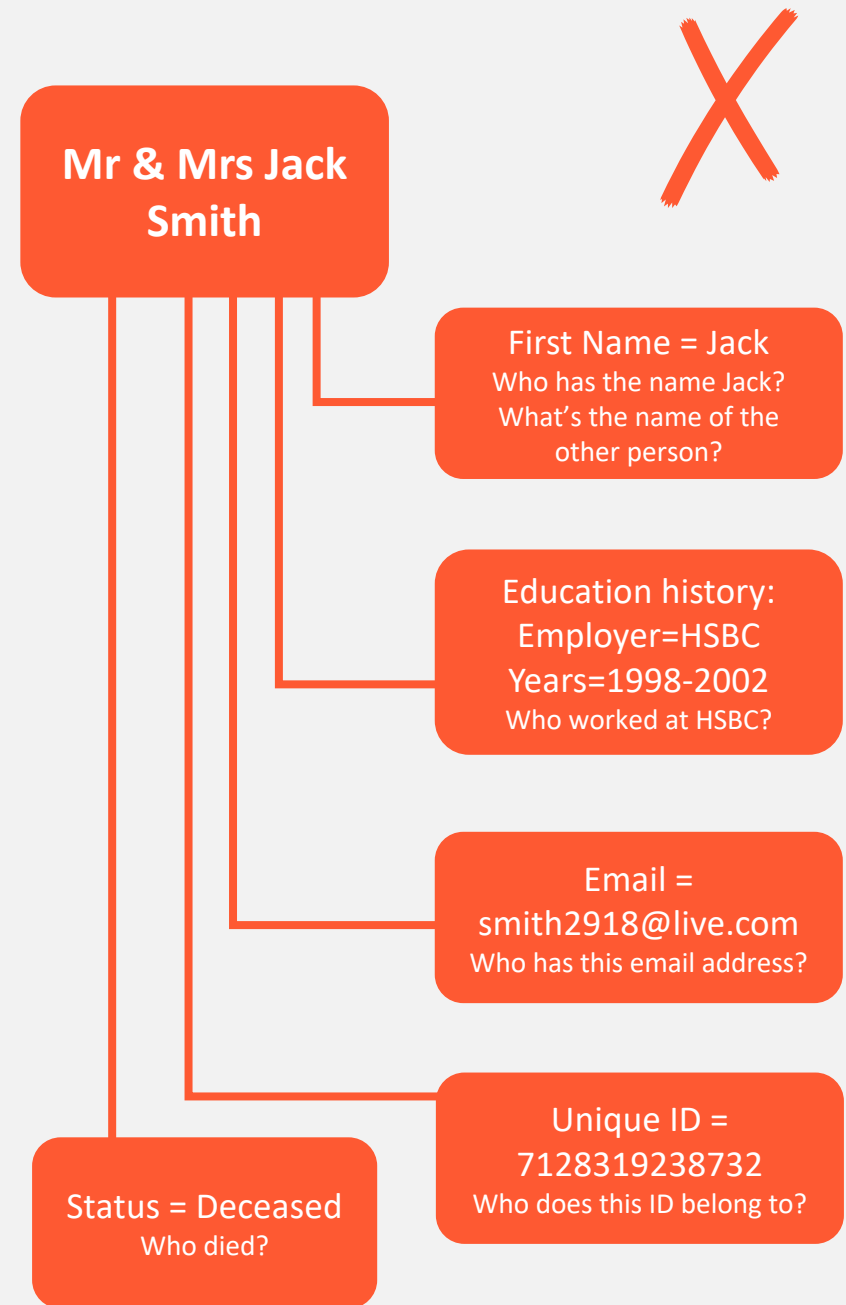
Mathematics teacher for two years, then taught Chemistry  
I was the old French teacher between 1987-1991  
Physics and English teacher  
I taught history for a short period of time then moved on to teaching Physics  
Mathematics 1 yr, French 1 yr  
I taught a lot of subjects!  
English Literature was my main subject, but I also stood in as a substitute biology teacher...  
French was my primary teaching subject  
I taught Biology then Chemistry  
Music teacher for 10 years and counting!  
Teach German, Italian and Frnech

Etc etc etc etc .....

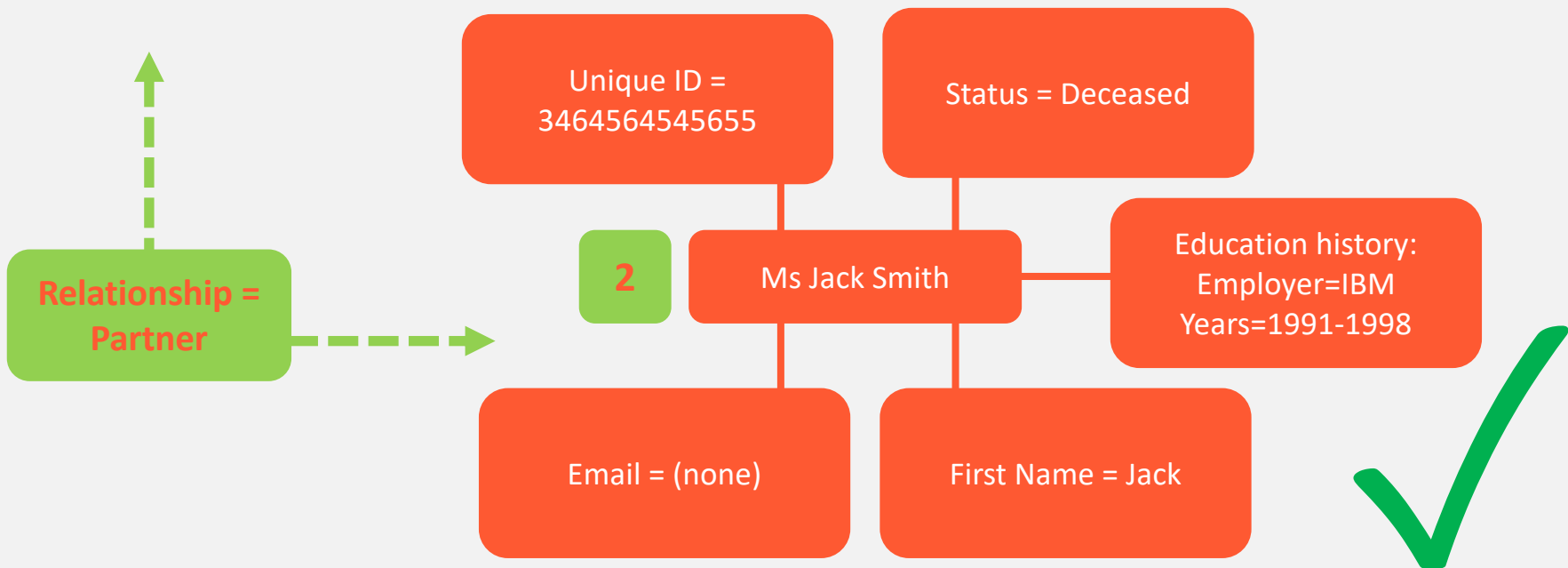
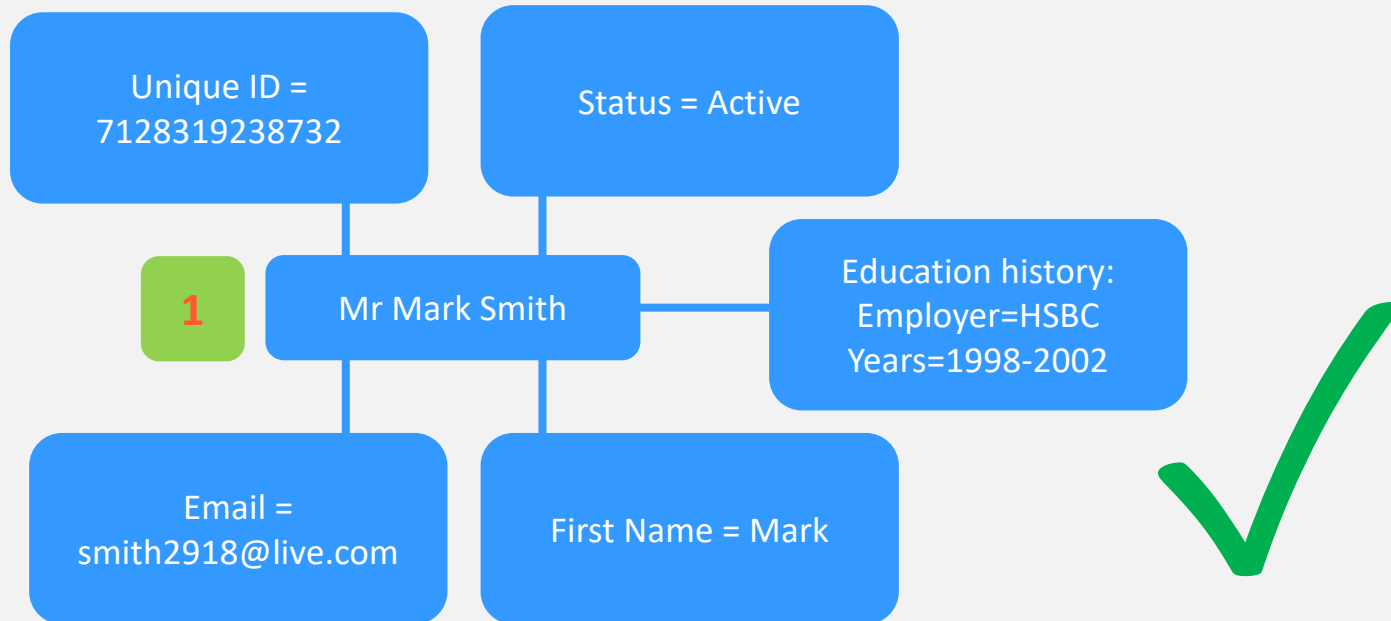


# Rule 3: One record per person

- ★ Referencing two (or more) people in one of your database records might be a short-term way to quickly add more data to your database, but it will cause complications and affect the quality of your data. We refer to these as 'joint records'
- ★ It will make it unclear which part of the record belongs to which person? (e.g. education history, work history etc.)
- ★ You no longer have 1 unique ID per person – two people share the same unique ID (see rule 1!)

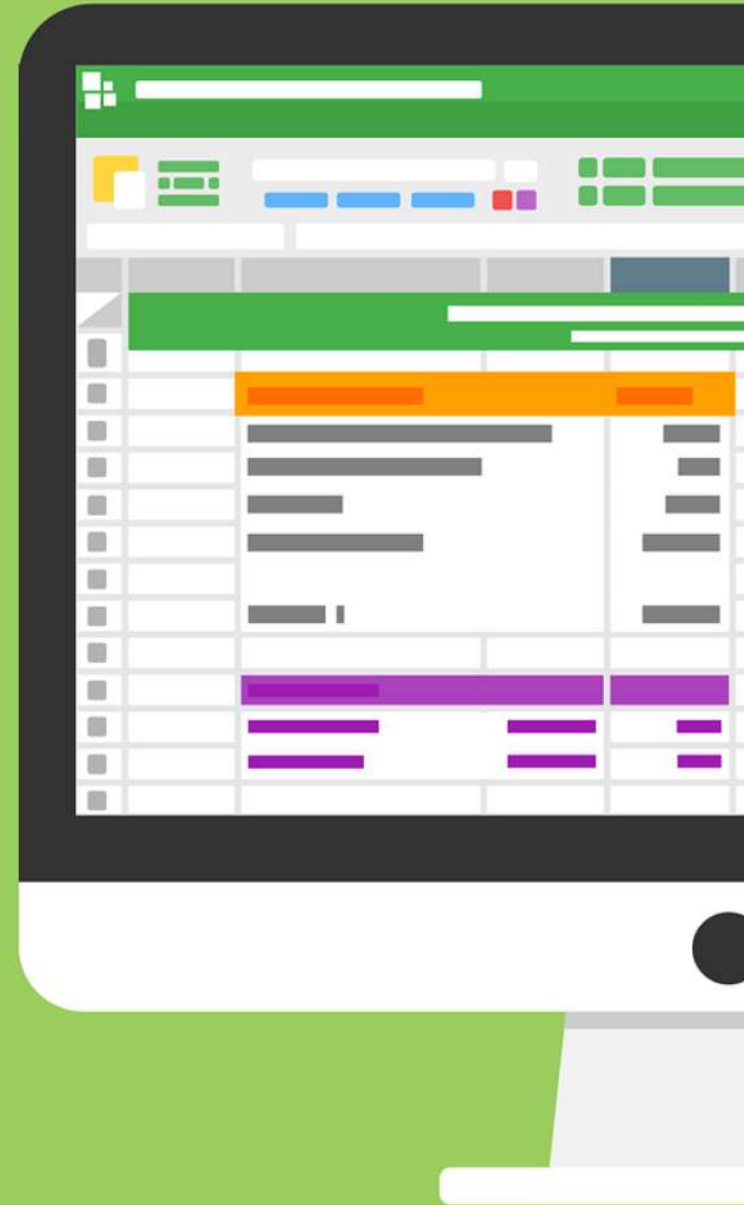




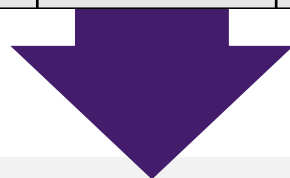


## Rule 4: Right information in the right field

- ★ **Date must match the description/purposes of the field**
- ★ Entering the **wrong data into the wrong field leads to inaccuracies in your entire dataset**. When running searches/reports, data issues like these will affect the output – “Oh I thought we had more alumni living in London!”
- ★ **Fixed lists (rule 2)** can help, limiting what a user can and can't enter into a certain field to prevent bad data from being incorrectly typed in but, ultimately, the best way to adhere to this rules to know your database well and ensure your colleagues do too!



Unique ID	Address Line 1	Address Line 2	Address Line 3	City	County/State	Country	Postcode
32748234789	15 Tyrell Street			Stoke Newington	London	United Kingdom	N16 1AA
33347239823	The Old Well	Haslemere		Waverley	Surrey	United Kingdom	GU11 8AA
89589489239	Flat 1, 5A Birch Crescent		Stoke Newington	London		UK	N16 1BB
81273828112	2918 Main st.			New York		United States of America	71829



Run database search for City="London".

**1 Result**

*"That's not right, I definitely have two Londoners in my database!"*

Run database search for County/State="New York".

**No Results**

*"Why hasn't Mr Smith come up in my results? He definitely lives in New York..."*

Run database search for Country="United Kingdom"

**2 Results**

*"There's something wrong, we have an alum in Birch Crescent but she's not on our results list"*



Unique ID	Address Line 1	Address Line 2	Address Line 3	City	County/State	Country	Postcode
32748234789	15 Tyrell Street		Stoke Newington	London		United Kingdom	N16 1AA
33347239823	The Old Well	Haslemere		Waverley	Surrey	United Kingdom	GU11 8AA
89589489239	Flat 1, 5A Birch Crescent		Stoke Newington	London		United Kingdom	N16 1BB
81273828112	2918 Main st.			New York	New York	United States of America	71829



Run database search for City="London".

**2 Results**

Run database search for County/State="New York".

**1 Result**

Run database search for Country="United Kingdom".

**3 Results**



# Rule 5: Keep it consistent!

- ★ **Establish/ set up data entry guidelines** and standards for your institution
- ★ If **capturing new data**, think about what answers you would need to see and how you can use them
- ★ If you are **cleaning data**, look at all the different variations and give them a new consistent term
- ★ **Fixed lists and data validation** will help you in doing this, only allowing a user to enter/select a pre-approved item or value

User groups
Former Student
Alumni
Ex-Student
Alumni
Former Student
Former Student
Past Student



User groups
Former Student
Former Student
Former Student
Former Student
Former Student
Former Student
Former Student
Former Student



# Rule 6: Remove your duplicates

- ★ If you see a **duplicate in your database it's time to take action** – they are like pests! For every one you've spotted there may be a dozen more lurking
- ★ **Some database tools may give you the functionality of flagging up potential duplications** based on email address, unique IDs, names – use them wisely to understand and assess the scale of the problem!
- ★ **Once you've located your duplicate records, take action** by merging the affected profiles together, ensuring that the same person doesn't have 2+ records



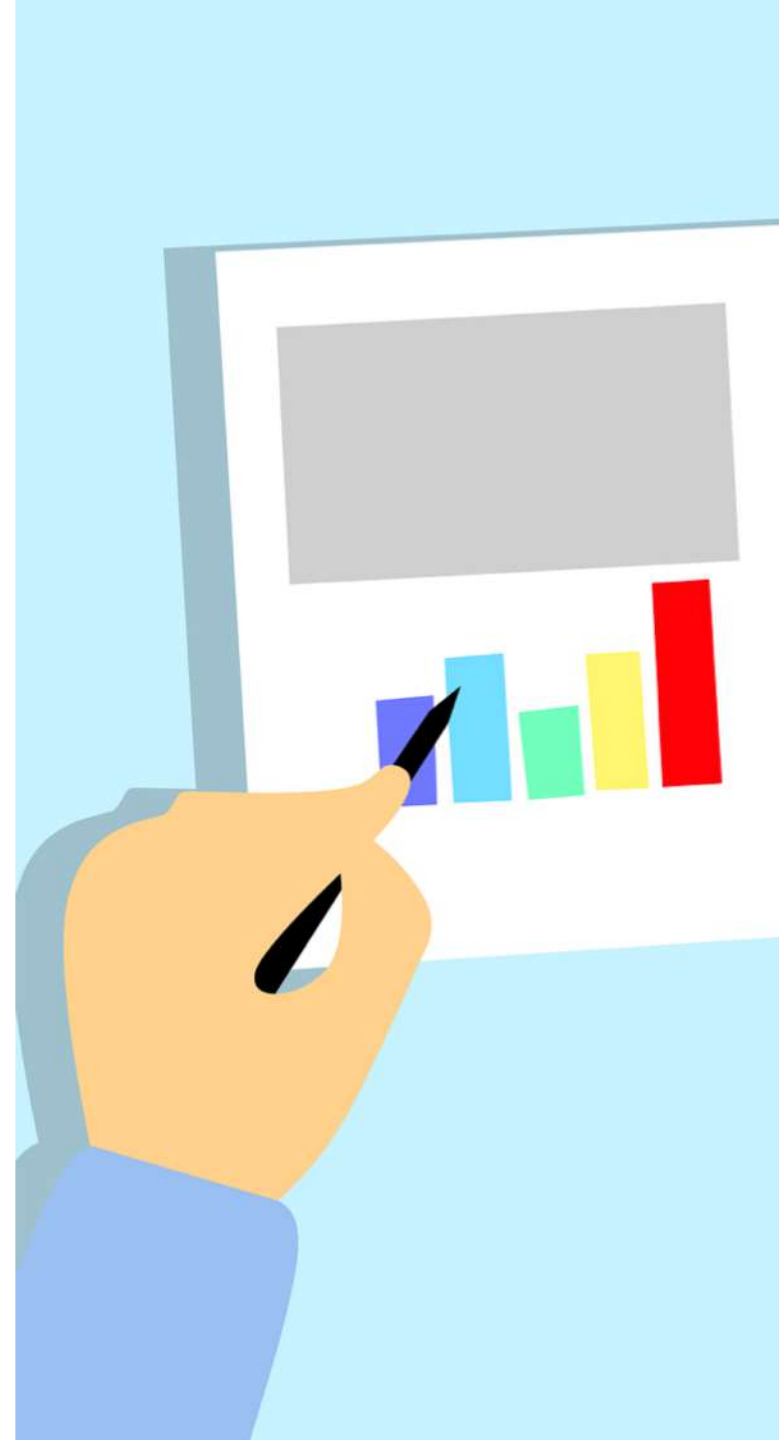
# Rule 7: Don't capture or hold data for the sake of it

When adding new data for your database, or reviewing old legacy data ask yourself why you want to capture or keep it?

- ★ *Does it serve a purpose today - purpose in 3 years time?*
- ★ *Why do I need this data - and what am I actively using it for?*
- ★ *How does this data help me and my institution?*
- ★ *What value does this data hold to me?*
- ★ *How accurate is the data I hold (quality and reputability)?*

If you can't clearly answer the questions, perhaps you should consider whether your data is relevant, a key principle of data storage under the GDPR – and reconsider whether you should proceed with recording the data, or whether you should consider clearing out old legacy data.

In a nutshell, personal data should not be retained longer than necessary, in relation to the purpose for which such data is processed.



# Summary



**Rule 1:** Always use Unique IDs



**Rule 2:** Validation is your friend



**Rule 3:** One record per person



**Rule 4:** Right information in the right field



**Rule 5:** Keep it consistent



**Rule 6:** Remove your duplicates



**Rule 7:** Don't capture or hold data for the sake of it



# About ToucanTech



ToucanTech is an easy, affordable and beautiful system to securely manage your alumni and supporter records, email newsletters, online payments, events, mentoring and more! Used by 100+ schools, with 100% renewal rate, we're passionate about clean data and engaged communities.





# Thank you!

[www.toucantech.com](http://www.toucantech.com) | [hello@toucantech.com](mailto:hello@toucantech.com) | +44 (0) 203 189 1774