

Collecting information with care

Opal Symes – she/her
@software_opal

The problem

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

3

We as technologists build systems that people interact with every day; and a lot of the time we go into those systems without thinking about the assumptions we have about them. And without questioning if those assumptions are valid. So today we're going to walk through some examples and challenge some assumptions about the way we should build these things.

At the end I hope that you'll have a better understanding of the assumptions we hold, how to challenge them, and how to build inclusive systems.

Be aware of my assumptions too

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

4

If your piece of technology needs to be culturally sensitive, especially around things like race, indigenous status; I am absolutely not the right person for the job. I'll put links at the end to some talks by people who are better suited to talking about these things.

Also, if you are a company needing to understand this you should look at hiring people who are well versed in these areas, hopefully by way of being in those marginalised groups, and doing so early in the process.

About Me

DevOps

Elixir, JavaScript, Python, and Rust

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

5

I'm a Developer and operations ... scientist working at a small data platform start-up based in Wellington. I write mostly Elixir and JavaScript, with a healthy side of Terraform. I also write Python and Rust in my spare time.

It's really interesting to build an extensible platform that is able to be used by many people whilst also being able to affect the way our clients use their data.

But this talk is not endorsed by them, I'm here in my own capacity.

About Me



COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

"Progress" Pride Flag by Daniel Quasar is licensed under CC-BY-NC-SA-4.0

I'm also a transgender woman.

I will happily admit that coming out gets easier; but it is never *that* easy.

It's always a game of chance, and in situations like this where I'm already out it's in my bio it's a lot easier.

But in other cases it's a matter of weighing up the options, figuring out what question they really want the answer to and finally figuring out which answer I feel safest giving.

Which was why I gave a talk in PyConAU earlier this year, and why I'm here. I want to show you why some questions are difficult to answer; and how to ask them better.

I'm going to talk a little bit about it later on in the talk.

Outline

Names

Emails

Gender

Changes & Deletion

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

7

This talk is broken down into 4 major parts.

Firstly we're going to talk about names; then we'll talk about emails and all the fun of validating them;

Then we'll cover gender, pronouns and titles.

Then finally we'll talk about changing these things and handling their deletion.

Names

So, lets begin with names and email addresses. For something quite simple; there's a lot that can go wrong.

Names

First Name*

Opal

Last Name*

Symes

Dear **Opal** **Symes**,

We're just saying hi!

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

9

So, lets start off with something simple, we'll collect the users first and last names and then when we come to use it, we'll just concatenate them together to create this sample email. And to make sure that users enter something we'll make both fields required.

So I'd put in Opal, and Symes, and when that gets rendered we get "Dear Opal Symes". Awesome.

But, those required fields are going to cause problems for people without a first name, or without a last name, so the simple fix is to change the form to require something in at least one of the fields.

Names

Required, but not really

First Name*
Opal

Last Name*
Symes

Dear Opal Symes,
We're just saying hi!

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

10

Awesome, so we've gotten rid of those pesky required fields, and now we just make sure there's something in at least one of them; and we're now able to handle users without a first name; or without a last name. So let's tackle the next problem ...

Names

Name

Clíodhna

We think you've mis-typed your first name - please try again

Ní

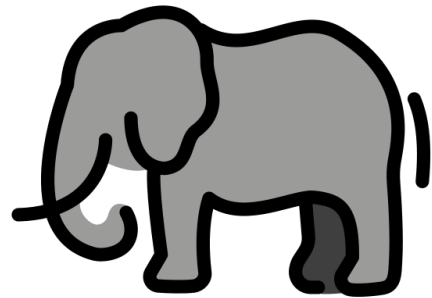
COLLECTING INFORMATION WITH CARE - @SOFTWARE_OPAL

11

Ah. Err ... I don't really know how to break it to this website ... but that message is really bad. As is the restriction on the characters in the name. It seems like the form isn't liking the accent above the 'i'. Which means we need to talk about the elephant in the room ...

Unicode

“Unicode covers all the characters for all the writing systems of the world, modern and ancient.”



Emoji designed by OpenMoji CC BY-SA 4.0

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

12

... sorry, that joke was just too good to not include. I'm of course talking about Unicode.

So, Unicode ... Unicode aims to be able to store all the characters for all the languages in the world, and it's doing a pretty good job of this. There's Egyptian hieroglyphics, Cyrillic and Greek characters. But the thing that really has made everyone take notice is Emoji.

This means even people with the privilege of having ASCII encodable names have needed to care about the way characters are encoded.

And there are a few gotchas, and I'm only going to scratch the surface; but I would strongly suggest using Emoji to check for encoding problems.

My favourite are emails and print, which have a tendency to mangle even with the best intentions. But first we must store them somewhere ...

Databases

Check production supports Emoji ... everywhere

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

13

Which brings us to databases. Now I don't have a lot of experience with building suitable database schemas; but the experience I have had with databases has been ... fraught. A previous project of mine fell over when one of our users logged in because we hadn't set up MySQL correctly. Which is why I've dedicated some slides to it. The easiest way for me to test was to put Emoji in all the text fields. You'll really easily find all the places that are mangling them and hopefully be able to trace them and fix them.

But the very first problem was getting the Emoji in the database ...

MySQL Character Encodings

utf8mb3 – Added to MySQL 4.1 in 2004 to cover “almost all modern languages”

utf8mb4 – Added to MySQL 5.5 in 2010 to cover the rest of the Unicode spec

utf8 = utf8mb4 – Default switched in MySQL 8(maybe)

<https://mysqlserverteam.com/mysql-8-0-when-to-use-utf8mb3-over-utf8mb4>

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

14

and this was because we were using MySQL 5 with the default character encoding of utf8mb3; which despite the name isn't utf8; it's more utf24. And if you're wondering "Opal; utf24 doesn't sound like it's standards compliant"; and you'd be right. For historical reasons; mostly around size on disk; MySQL decided to implement Unicode using only the 3 low bytes which was mostly fine back then.

Since then; they have introduced utf8mb4; which is technically utf32. But this encoding can hold all of the Unicode characters rather than a subset of them.

And then in MySQL 8, they have switched the utf8 character encoding to the 4 byte version and I believe that it is the default encoding for new databases.

MariaDB is ... different; despite being based on MySQL and maintaining feature parity they haven't switched the default encoding for new databases. Unless you're using Debian, which has.

Database Tweaks

Set the character set & collation on each database, table & text column

Set the defaults so new columns are correct

Set connection options so things don't get lost in the transfer

<https://mathiasbynens.be/notes/mysql-utf8mb4>

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

15

So, now we know what the encodings are and why you should definitely only use utf8mb4. To migrate you will need to change the database, table and column. And you also need to make sure you're connecting to the database using the same encoding to prevent things being dropped at that level. And finally you need to test. Preferably with characters that you can easily tell if things are going wrong. As I mentioned emoji is great for this.

It is still perfectly possible to mess up the character encoding in other databases and systems; but this one happens to be one I ran into last year and I expect it to still catch people out this decade.

Names – Recap

First Name

Last Name

First Name*

Opal

Last Name*

Symes

Dear **Opal Symes**,

We're just saying hi!

<https://www.w3.org/International/questions/qa-personal-names>

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

16

Phew, we survived our brief encounter with Unicode. So a refresher on where we are. We have a form with a first and last name; and we're going to display them by concatenating them. We also handled the case where a user doesn't have both a first and last name. So this is ok; but we need to talk about this form when working with other languages and cultures.

I'm specifically going to use Chinese as an example which I have shamelessly taken from the W3C's brilliant document on names in different cultures.

Names

First Name

毛

Last Name

泽覃

Dear 毛 泽覃,

We're just saying hi!

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

17

In this example we're looking at someone called Mao Ze Tan. But unlike in English where the first name is the given name. In Chinese the first part of the name is the family name, the middle part is the generational name; and the last part is the given name.

So, whilst our existing system works; by making the user split their name; and re-joining it. This could lead to problems.

Names

Switching to displaying first name only

Dear **☹**,
We're just saying hi!

Dear **Symes**,
We're just saying hi!

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

18

This splitting is not a huge problem right now; but if we were to change the email template to use just the first name it would look super weird. It would be equivalent to saying Dear Symes. This is less than ideal. Before we can improve this, we should look at what makes a name

Names in depth

Names are wide and varied across cultures

The first 'word' in a name is not necessarily someone's given name

Names can be long

<https://www.w3.org/International/questions/qa-personal-names>

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

19

There is a huge variety in the way names are written and used in cultures around the world. One important thing is that the first word in a name isn't necessarily someone's given name. The

If you would like to learn more about the different schematics of names you should look at the W3 article on names. There are heaps of examples and I highly recommend that you look at them and try using them in your systems to challenge the system's assumptions.

At the end of the day, what they are trying to guide you towards, and what I too am trying to guide you towards is a better system.

Names

Full name & Greeting

Full Name*

Opal Symes

Greeting*

Opal


COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

20

So what we should do is just use a greeting field. And if we need the full name, we just use a full name field.

Names

Multiple names



Full Name*

Symes

Greeting*

Opal

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

21

Now this setup can be used together to allow people with multiple names to use your system; by allowing them to put their full or legal name, by the way if that distinction is important you should definitely call it out, and then their preferred name in the greeting.

Names – Summary

Full name/Greeting

Unicode all the way down

No length limits

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

22

Awesome, we've created a form that works across cultures. And we're allowing people to enter their name and we can store it by making sure our database supports the entire Unicode range. We're also not limiting the length of that field so that users with long names can enter them completely.

Email

Phew, we've gotten over the complicated and convoluted bit of the talk and now onto emails. Ok, I lied; emails are also complicated and convoluted

Guess the invalid email

français@opal.codes

opal.is(awesome)@opal.codes

"opal is doing ... good"@opal.codes

OPAL.talks+stuff@opal.codes

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

24

So I figured we'd play a quick game of which of these emails aren't valid using the most recent RFC.

Guess the invalid email

français@opal.codes

- Valid with RFC6530

opal.is(awesome)@opal.codes

Comments '()' at the start or end

"opal is doing ... good"@opal.codes

Spaces are allowed in quotes

OPAL.talks+stuff@opal.codes

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

25

Well, they are all in fact valid email addresses.

The first one is valid because RFC6530 allows for non-ascii email addresses.

The second one uses a comment, the bit in parenthesis

The third one is valid because you can use quotes in the local part to enclose spaces, and to allow dots at the end; but dots at the end are not allowed outside of quotes

And the fourth one uses case as well as dots and plus symbols.

That being said, my email provider doesn't let me use the top 3; so ... maybe they're not valid ...

Emails are hard

Too many websites reject this

Email*

opal@opal.codes

https://en.wikipedia.org/wiki/Email_address#Examples

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

26

No matter if they are valid or not, there are a whole heap of rules around what makes a valid email address; and I would strongly suggest that you don't try and implement something yourself.

If you want some more fun the Wikipedia article below has some great examples.

But we should still make sure the emails are valid ...

Validating emails

Does it have an '@'

Similarity to well known email providers:

Think gnail vs gmail

Will it be accepted by your email sender?

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

27

The easiest form of validation you can do is check to see if the email has an @ symbol, and a dot afterwards. This will catch some invalid email addresses; the final check will be does your email system support sending it

Other things you can do to try and catch common errors is to see if the bit after the @ looks like a common email provider. Someone may type gnail instead of gmail and it'd be good to catch it on the UI; but don't prevent users from submitting emails if they're close; but not a match.

But if you want to get really fancy you can ...

Validating emails

Look for MX records

Ask the SMTP server if the username is valid

On the more complex end you can start doing actual email parsing. You can also do cool things like checking if the domain name has MX records, and you could even connect to the SMTP server and ask them if a username exists.

In my opinion this is overkill; but your situation may be different.

But there are libraries that can help with this.

Emails

Simple checks to catch common errors

Don't roll your own checking; use your mail server, or mail library to make sure you can send it.

So what have we learnt,

Names should be treated as a single field unless you have a specific need to split them. Use UTF-8 to store them and make sure that they accept the full range of characters, and that they render correctly throughout your system. Finally don't limit the length of name fields, either in the form or in the database.

With email, you should do simple checks that catch common mis-typings but don't prevent people from submitting the 'mistyped' email. There are ways you can do complicated checking but the Regular expressions frequently miss things. Stick to well defined email parsing libraries if you have to. Also don't try to limit an email's length.

Gender

Ok, we're over half way through the talk and I think the hard stuff is behind us. Now we just need to talk about gender.

Transgender & Non-binary

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

31

Before I can talk about gender, I need to get some definitions out of the way.



Transgender

Transgender

– Different gender to the one assigned at birth

Cisgender

– Same gender as the one assigned at birth

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL



32



Non-binary

Gender that isn't male or female



COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

Gender

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

34

Gender

What is your gender?

Male

Female

So, we will start of with a form like this one. I see this all the time on websites, even airlines; which I'll cover in a moment. But the problem with this form is that it does not list all the options

Gender

What is your favourite colour?

- Pink
- Purple

It would be the equivalent of having a question asking for someone's favourite colour and it only listed two colours.

Gender – there are more than two

What is your gender?

- Male
- Female
- Non-binary

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

37

So, we know there are more than two genders, and so you could do this. This is mostly ok. It's miles ahead of the previous one. But some people take this to mean 'other'.

Gender – Alternative

Gender

Transgender Woman

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

38

Alternatively you could let people enter their gender in a text box. This is great if you want to understand the composition of your community; You will have seen this when you signed up for the conference tickets.

This will allow your users to choose how much information to provide. So for me I put in transgender woman. But I may have chosen to not reveal my 'transness' and just put woman.

One drawback with this is that you will need to do some postprocessing of the data to be able to get some statistics out of it.

Gender – Alternative

Open Demographics

<http://nikkistevens.com/open-demographics/>

Do you consider yourself to be transgender?
 Yes
 No
 Questioning

Do you consider yourself to be gender non-conforming, gender diverse, genderqueer?
 Yes
 No
 Questioning

Are you intersex?
 Yes
 No
 I don't know

Where do you identify on the gender spectrum (check all that apply)?
 Woman
 Demi-girl
 Man
 Demi-boy
 Non-binary
 Demi-non-binary
 Genderqueer
 Genderflux
 Genderfluid

... snip ...

Neutrois
 Androgynous
 Androgynous
 Prefer not to answer
 Self Identify: _____

If you want to avoid the extra work of post-processing your data you could look at using this form that was built by OpenDemographics. They have a heap of sample forms on all sorts of things and I would highly recommend looking through them if you need to get demographic information.

Community composition

Please indicate any of the following underrepresented communities that you identify with. We have provided some examples of each, but these are not exhaustive—if you feel you identify, please mark the box.

This field will remain private, and will only be used to provide anonymized, aggregate data about the makeup of the Drupal community.

- Ability
- Age
- Ethnicity
- Gender
- Gender identity
- Race
- Religion
- Sexual orientation
- Socio-economic status/class
- Learning differences
- Family composition
- None
- Prefer not to answer

We have chosen these categories based on the [Big 8](#), which is a list of cultural identifiers. [Learn more about why we collect this type of data.](#)

But maybe you don't actually care about the gender; but about the composition of your users; and you could look at this sort of form which simply asks if you identify as a member of a minority.

This was taken directly from Drupal.org

Pronouns

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

41

So, we've collected the users gender and some of you may be wondering if we should use that information to change the pronouns we refer to the user.

Gender does NOT
imply pronouns

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

42

In short ... No.

Pronouns

What pronouns should we use to refer to you?

- She
- They
- He

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

43

A user's gender does not imply that they use a particular set of pronouns. So, if you are going to add this you need to have these three. And you need to use them everywhere.

Pronouns - Translations

Some languages lack a gender-neutral 'pronoun'

Some languages lack gendered 'pronouns'

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

44

The thing to watch for when you are translating these pronouns is for languages which do not have gender-neutral pronouns. You should talk to your translators about this, and if you're unable to resolve it then maybe provide the users an option for their preferred gendered pronoun.

The other way around, for languages that don't have gendered pronouns; then you should ... just use the ungendered one.

Pronoun Island(pronoun.is)

pronoun.is is a website for personal pronoun usage examples

here are some pronouns the site knows about:

- [she](#)
- [he](#)
- [they/.../themselves](#)
- [ze/hir](#)
- [ze/zir](#)
- [xey/xem/xyr](#)

[see all pronouns in the database](#)

Here are some example sentences using my she/her pronouns:

She went to the park.

I went with **her**.

She brought **her** frisbee.

At least I think it was **hers**.

She threw the frisbee to **herself**.

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

45

If you're not really worrying about translations; then I would suggest looking at opening up other pronoun options beyond the three. There is a great website called pronoun.is, and it has a whole heap of them along with the variants and examples of their usage.

They have a

Titles

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

46

Gender does NOT
imply titles

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

47

In short ... No.

Titles

- Title
- None
 - Mx
 - Mrs
 - Ms
 - Miss
 - Mr

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

48

So, titles are not defined by gender, and you should provide at the very least the option of “Mx” and not having a title. Alternatively you could save yourself the hassle and not use one. Or let people type their own.

Clothing preferences

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

49

Gender does NOT
imply clothing
preferences

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

50

In short ... No.

Preferences

I want to see:

- Femme shoes
- Masc shoes
- Dresses
- Suits
- Ballgowns

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

51

As we have seen before, gender is not useful for defining what someone wants to be called, what title they use and what they want to wear; so if you are asking for someone's gender to show them specific products; maybe don't. Use a list of interests. And there are a heap of reasons to do this; the most ... obvious ... is shopping for someone else. But you may not like skirts, or makeup, or tall shoes.

Gender 2

The compliance reconing

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

52

WGEA

The Act defines gender as **female or male**. If an employee does not identify as either female or male, you do not need to include them in your workplace profile.

Ah yes, WGEA. So, some background for those who don't really know what it is. The Workplace Gender Equality Agency is an Australian government agency that's tasked with promoting gender equality in Australian workplaces.

One of the ways it does this is by requiring companies over a certain size to report the gender composition of their workforce ... but ...

Right here on page 14 of the reference guide they say this:

The Act defines gender as female or male.

Well ... we know that gender isn't the simple binary; but they have to work within the act and so do employers.

So ... how can we work around this. Well, from what I understand it's a declaration so one could architect a form that allows non-binary people to say "Count me as this gender". How you want to do this will depend on what message you want to send; and also what people will feel comfortable answering with

WGEA

Gender for the purposes of WGEA reporting

- Male
- Female
- Do not include me in the report

 Due to the limitations in the WGEA reporting guidelines we are required to either report you as having a binary gender, or not include you in the report.

And here's my entry

WGEA & Demographics

What is your gender?

- Male
- Female
- Neither male or female

ⓘ The selection below will only be used for compiling the WGEA report.

Gender for the purposes of WGEA reporting

- Male
- Female
- Do not include me in the report

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

55

Alternatively if you are collecting some extra demographic information instead of just the gender for WGEA, you could split the question up into two.; and then when you select the non-binary option, you get the previous form.

Gender – Passports

Gender(as shown on your passport)

- M – Male
- F – Female
- X – Other

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

56

If you are asking for passport information, then you need to support at the very least the 3 shown here. And before you ask, no, the airlines still don't. But you could be the first. The fourth maybe option was being suggested by the UK government which was to remove it entirely. And so you may need that as an option in the future.

Gender – Summary

Gender is not a binary; don't use a binary to capture it

Gender is not a replacement for other questions like preferences

Give people reasons why you have to capture a binary

Gender APIs - Don't

gender api

All News Images Shopping Videos More Settings Tools

About 102,000,000 results (0.42 seconds)

Gender API - Determines the gender of a first name
<https://gender-api.com>
Keep your registration form simple. Optimize your conversations and let us determine the **gender** of your customer.
Faq · Gender-API API Clients · Api Docs · Gender-API.com account

Faq - Gender API - Determines the gender of a first name
<https://gender-api.com/frequently-asked-questions>
The new European General Data Protection Regulation (GDPR) goes into effect on May 25, 2018. We can assure you that all **Gender-API.com** services will be in ...

Gender API
<https://genderapi.io>
Detection **gender** from name or email address. And also you can find social media users **gender's** by their username. Find your name is female or male, get ...
Instagram Gender · İsimden Cinsiyet · FAQ · Name Geschlecht

Genderize.io | Determine the gender of a name
<https://genderize.io>
Free API to predict the **gender** of a name. Easy integration into any platform. Over 250,000 names in our database. Serving more than 1 million requests/day.
Localization · Rate Limiting · Responses & Errors

COLLECTING INFORMATION WITH CARE - @SOFTWARE_OPAL

58

Some people may be tempted to skip all this work and just use something like these Gender APIs, but these are really really shitty. They rely on things that aren't gender to figure out gender. Please, don't use them.

Data retention

So, to finish all this off, lets talk about what to do with all this data once you've collected it. Now I'm not an expert on GDPR or the CCPA, so take what I say with a grain of salt; but what I'm going to talk about is more general than that. Specifically around updating and deleting data.

Changing

Make it simple to change

Give users options if it isn't

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

60

How many people have accidentally typed their name in wrong on a website only to find that they mistyped their name. I know I have ... so many times ... with my deadname, and my title. And you would think it would be easy to change them, and for some websites it's super easy. Other times it's like pulling teeth.

I have had an experience with a rewards program where I wanted to change my title; possibly the most useless part of a name; and I needed to call them up and then send in a copy of my ID. Needless to say I'm no longer a member of that particular rewards program.

So ... what can we do. Well we can make it easy to change the information. And when you can't make it easy to change, you should give people the option to have their preferred name appear places where it doesn't matter.

Changing Name

Personal details

If your name has changed, or is incorrect, please visit our [Help & Contact](#) page to update it.

Name

Mrs [REDACTED]

Middle name optional

Preferred name optional

Display my preferred name

We'll display this name on the seatback entertainment screen, and other communications when we can. Don't have one? We'll use your first name instead.

Date of birth

COLLECTING INFORMATION WITH CARE - @SOFTWARE_OPAL

61

Going back to another rewards program I'm a part of, which still has my deadname on it. And this is probably the best option I've seen. As I mentioned earlier, my deadname is still on my passport, and so having the option of not being deadnamed by emails is really nice. As an aside, I have no idea why middle-name is not part of the 'name' but ... that will probably remain one of the unsolved mysteries of the universe.

Changing username

Username are often names

Changing them comes with some caveats

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

62

With changing names, you also need to support changing usernames, because usernames are often a variant of their name, even mine.

It's important to think about this beyond just changing a column in a table. What will happen to services that connect to your systems; if they use the username your users could lose access to their accounts, or worse another user could gain access to someone else's account.

Changing usernames

Use IDs/UUIDs to link to users instead of usernames

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

63

The best way to deal with this is to avoid using the username as a shareable identifier. In my previous role we were wrestling with changing usernames in LDAP, and it was hard. One of the key reasons why was that there were so many different systems, and most of them used the username as the primary key.

The route I would suggest you take is migrate away from usernames; make it abundantly clear that a username is just like a name; it can change. Use IDs, if you want to get really fancy UUIDs.

Changing usernames

What happens to old usernames?

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

64

The final thing with changing usernames is to think about how to handle the old username. Places like GitHub will just switch the URLs over; and let you know on the CLI that the repo has changed; but don't do anything on the website. Visiting my old username gives a 404. This might be less than ideal; especially if there are a lot of links to your GitHub account.

You will probably want to look at preventing users registering with that old username; otherwise your users will probably do that for themselves.

Changing Email

There are security implications

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

65

So once you've accounted for changes to names; we should talk about emails again. This is more an observation, than anything else. Access to your email accounts will allow an attacker to gain access to your accounts by resetting your password. This means changing to emails could be used to hijack an account. Now, I'm not suggesting make it hard; trust me, that's just as much hassle. But what you should definitely think about when writing the process is this stress case.

My preference would be an email sent to both the old and new email address that is needed to confirm the change. Or just a confirmation email. Or anything. To be honest it would be better than the previous place.

Deleting data

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

66

And now we can talk about deletion. If you've made it really easy to sign up then you should, out of a courtesy to your users, make it as easy to delete the account; and all the data associated with it.

You'll want to confirm this; probably by email. And you may want to place a time delay on the deletion in order to protect users.

Conclusion

Names

Unicode everywhere

Don't assume first name is someone's given name

Make changing them easy

Use a "preferred" name when changing is hard

Emails

Follow the RFC(5322). Accept everything your outgoing service supports

Make changing email easy; but secure

Gender

Gender isn't a binary

Gender doesn't dictate anything else

Maybe ask a different question

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

70

Data retention

Make it easy to update information stored

Be aware of the risks around changing usernames & emails

Allow users to delete their data easily.

Things I didn't cover

Security

Privacy

Passwords

Encryption(in flight & at rest)

Wording & Phrasing

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

72

Thank you

Slides, notes & references – opal.codes/lca2020

COLLECTING INFORMATION WITH CARE – @SOFTWARE_OPAL

73

These slides are available at opal.codes/lca2020. And I would like to say a huge thank you to Kathy Reid for reviewing my talk.