

# Level 1: Stable HTTP URIs

**What is an HTTP URI?**

# What is an HTTP URI?



Uniform Resource Identifier

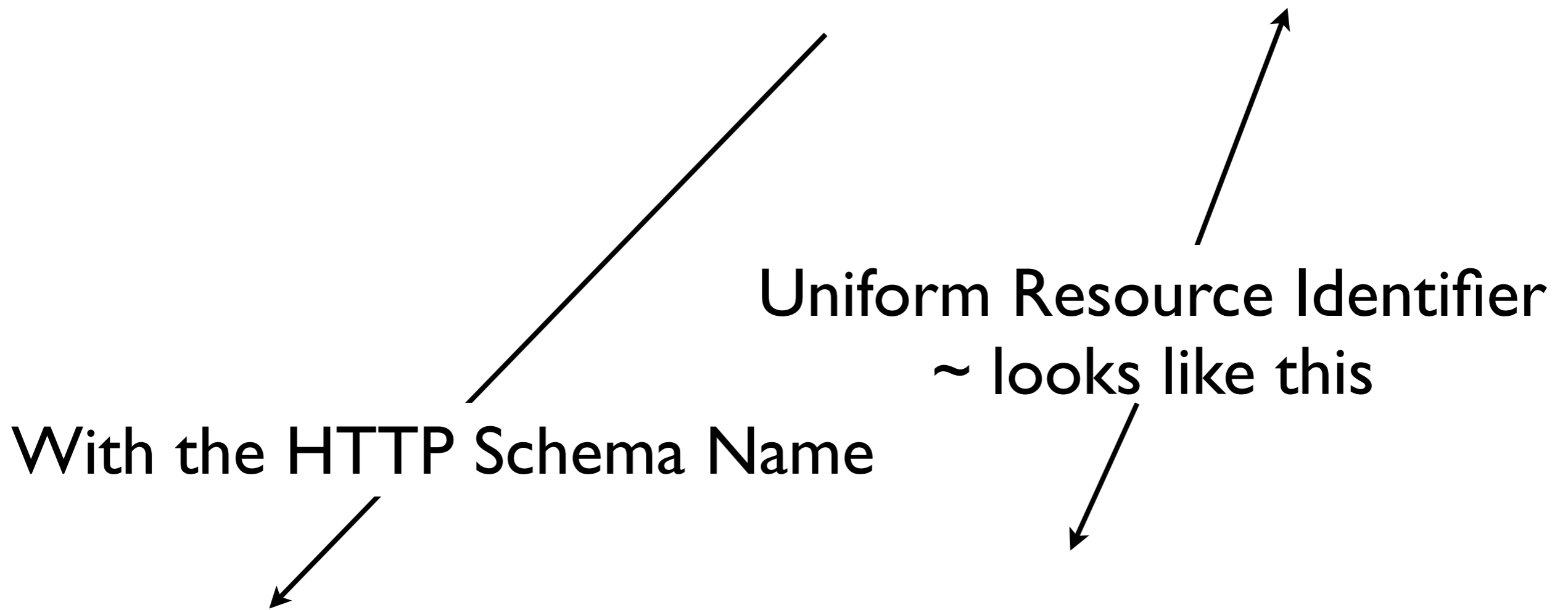
# What is an HTTP URI?

Uniform Resource Identifier  
~ looks like this



<scheme name> : <hierarchical part> [ ? <query> ] [ # <fragment> ]

# What is an HTTP URI?



<scheme name> : <hierarchical part> [ ? <query> ] [ # <fragment> ]

# What is an HTTP URI?

Uniform Resource Identifier  
~ looks like this

With the HTTP Schema Name

<scheme name> : <hierarchical part> [ ? <query> ] [ # <fragment> ]

HTTP implies a resolution mechanism so sometimes called a URL (location) - avoid but don't get a hang up!

# Names vs Locations

# Names vs Locations

- Dangerously near philosophy (or religion)



# Names vs Locations

- Dangerously near philosophy (or religion)
- Roger Hyam == a name

# Names vs Locations

- Dangerously near philosophy (or religion)
- Roger Hyam == a name
- More than one Roger Hyam and no authority to say which is the “real” one

# Names vs Locations

- Dangerously near philosophy (or religion)
- Roger Hyam == a name
- More than one Roger Hyam and no authority to say which is the “real” one
- Authority & trust only available when name combined with an address - a precise name?

# Names vs Locations

- Dangerously near philosophy (or religion)
- Roger Hyam == a name
- More than one Roger Hyam and no authority to say which is the “real” one
- Authority & trust only available when name combined with an address - a precise name?
- Our HTTP URIs use the DNS system to provide an authoritative, global namespace

<http://data.rbge.org.uk/herb/E00421509>

<http://data.rbge.org.uk/herb/E00421509>

- URL Schema Name

<http://data.rbge.org.uk/herb/E00421509>

- URL Schema Name
- By definition we are using HTTP so there is nothing to talk about here

<http://data.rbge.org.uk/herb/E00421509>

- URL Schema Name
- By definition we are using HTTP so there is nothing to talk about here
- Note it is lowercase



<http://data.rbge.org.uk/herb/E00421509>

- URL Schema Name
- By definition we are using HTTP so there is nothing to talk about here
- Note it is lowercase
- `https` is an entirely different protocol

<http://data.rbge.org.uk/herb/E00421509>

<http://data.rbge.org.uk/herb/E00421509>

- First part of path is domain name

<http://data.rbge.org.uk/herb/E00421509>

- First part of path is domain name
- Not a separate part of the URI schema just a convention

<http://data.rbge.org.uk/herb/E00421509>

- First part of path is domain name
- Not a separate part of the URI schema just a convention
- Case insensitive but by convention we should always use lower case

<http://data.rbge.org.uk/herb/E00421509>

- First part of path is domain name
- Not a separate part of the URI schema just a convention
- Case insensitive but by convention we should always use lower case
- Provides authority, **responsibility** and trust layer

<http://data.rbge.org.uk/herb/E00421509>

- First part of path is domain name
- Not a separate part of the URI schema just a convention
- Case insensitive but by convention we should always use lower case
- Provides authority, **responsibility** and trust layer
- Whoever controls the domain name publishes the canonical data for the specimen

**DNS not IP Address**



# DNS not IP Address

- Quick statement of the obvious

# DNS not IP Address

- Quick statement of the obvious
- It would be legal to have HTTP URIs that use IP addresses instead of DNS entries

# DNS not IP Address

- Quick statement of the obvious
- It would be legal to have HTTP URIs that use IP addresses instead of DNS entries
- `http://192.168.150.50/herb/E00421509`

# DNS not IP Address

- Quick statement of the obvious
- It would be legal to have HTTP URIs that use IP addresses instead of DNS entries
- <http://192.168.150.50/herb/E00421509>
- <http://data.rbge.org.uk/herb/E00421509>

# DNS not IP Address

- Quick statement of the obvious
- It would be legal to have HTTP URIs that use IP addresses instead of DNS entries
- <http://192.168.150.50/herb/E00421509>
- <http://data.rbge.org.uk/herb/E00421509>
- But don't do it.

# DNS not IP Address

- Quick statement of the obvious
- It would be legal to have HTTP URIs that use IP addresses instead of DNS entries
- <http://192.168.150.50/herb/E00421509>
- <http://data.rbge.org.uk/herb/E00421509>
- But don't do it.
- We mean HTTP URIs with Domain Names only!

<http://data.rbge.org.uk/herb/E00421509>

<http://data.rbge.org.uk/herb/E00421509>

- After the slash is the object identifier



<http://data.rbge.org.uk/herb/E00421509>

- After the slash is the object identifier
- Handled by the server

<http://data.rbge.org.uk/herb/E00421509>

- After the slash is the object identifier
- Handled by the server
- May or may not be treated as case sensitive by the server

<http://data.rbge.org.uk/herb/E00421509>

- After the slash is the object identifier
- Handled by the server
- May or may not be treated as case sensitive by the server
- Should be treated as case sensitive in the identifier

<http://data.rbge.org.uk/herb/E00421509>

- After the slash is the object identifier
- Handled by the server
- May or may not be treated as case sensitive by the server
- Should be treated as case sensitive in the identifier
- None of our business but may include strategy for internal namespaces

# Case Sensitivity Summary

# Case Sensitivity Summary

- Different parts of the URI syntax may or may not be 'officially' case sensitive

# Case Sensitivity Summary

- Different parts of the URI syntax may or may not be 'officially' case sensitive
- RDF and related technologies treat URIs as case sensitive

# Case Sensitivity Summary

- Different parts of the URI syntax may or may not be 'officially' case sensitive
- RDF and related technologies treat URIs as case sensitive
- For maximum interoperability we should treat the entire URI string as case sensitive



# Case Sensitivity Summary

- Different parts of the URI syntax may or may not be 'officially' case sensitive
- RDF and related technologies treat URIs as case sensitive
- For maximum interoperability we should treat the entire URI string as case sensitive
- For readability we should tend to lowercase where possible

# Strategies

# Strategies

- Opaque or Human Readable

# Strategies

- Opaque or Human Readable
- Choosing domains and sub-domains

# Strategies

- Opaque or Human Readable
- Choosing domains and sub-domains
- Choosing object identifiers

# Opacity Rocks

# Opacity Rocks

- The ID of an object is not predicated on its attributes

# Opacity Rocks

- The ID of an object is not predicated on its attributes
- Non-computer / Philosophy students find this difficult



# Opacity Rocks

- The ID of an object is not predicated on its attributes
- Non-computer / Philosophy students find this difficult
- Why is it not dependent on attribute values?

# Opacity Rocks

- The ID of an object is not predicated on its attributes
- Non-computer / Philosophy students find this difficult
- Why is it not dependent on attribute values?
- So we can talk about the values of attributes without changing the identity of the thing we are talking about

# Opacity Rocks

- The ID of an object is not predicated on its attributes
- Non-computer / Philosophy students find this difficult
- Why is it not dependent on attribute values?
- So we can talk about the values of attributes without changing the identity of the thing we are talking about
- Sawing off the branch we are sitting on

If opacity rocks how come your  
URIs have attributes in them?

If opacity rocks how come your  
URIs have attributes in them?

<http://data.rbge.org.uk/herb/E00421509>

If opacity rocks how come your  
URIs have attributes in them?

<http://data.rbge.org.uk/herb/E00421509>

Type of organisation



If opacity rocks how come your URIs have attributes in them?

<http://data.rbge.org.uk/herb/E00421509>

Acronym

The diagram consists of two arrows pointing upwards from the labels 'Acronym' and 'Type of organisation' to the 'rbge' part of the URI 'http://data.rbge.org.uk/herb/E00421509'. The arrow from 'Acronym' points to the 'rb' part, and the arrow from 'Type of organisation' points to the 'ge' part.

Type of organisation

If opacity rocks how come your URIs have attributes in them?

<http://data.rbge.org.uk/herb/E00421509>

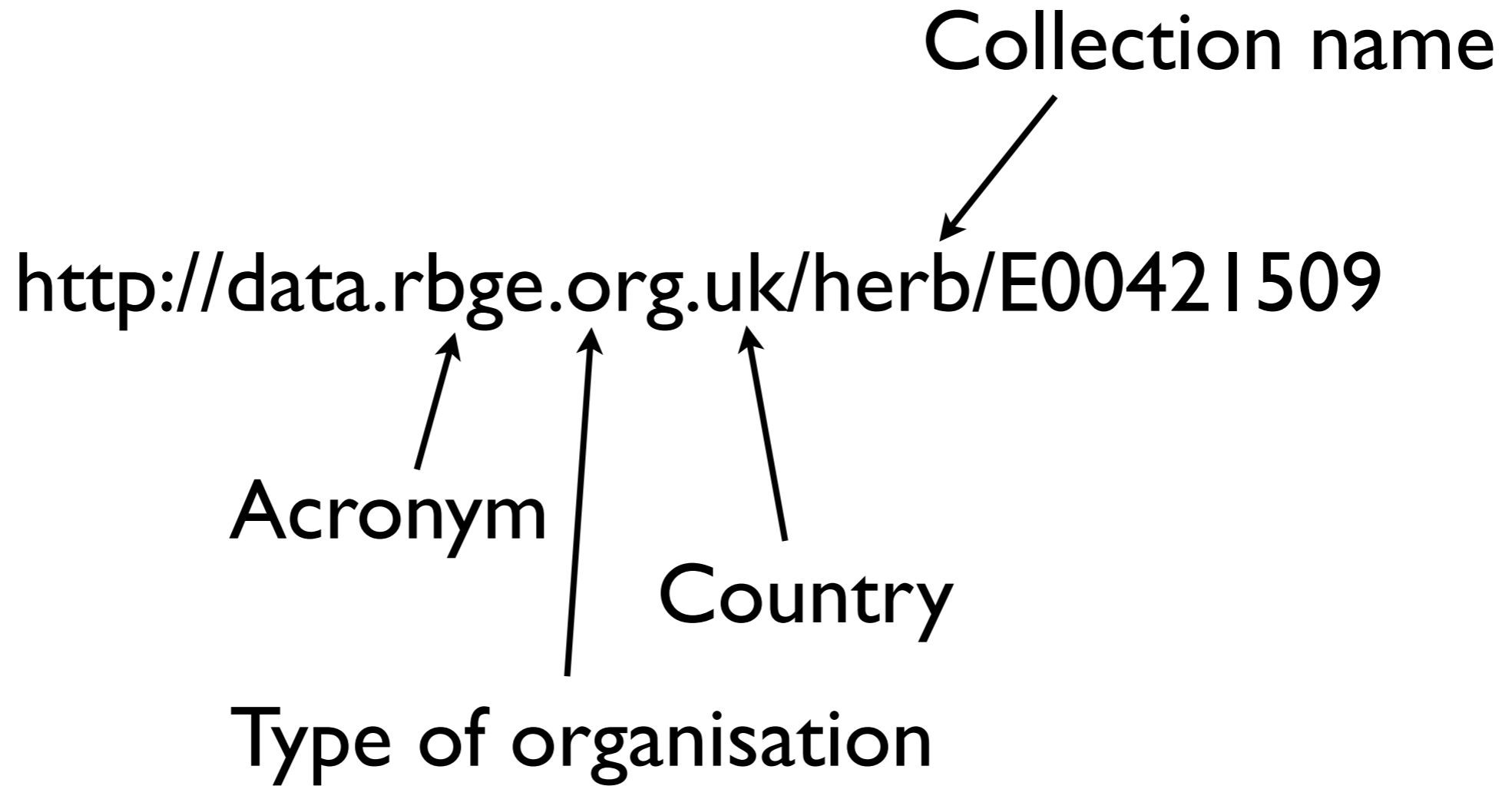
Acronym

Country

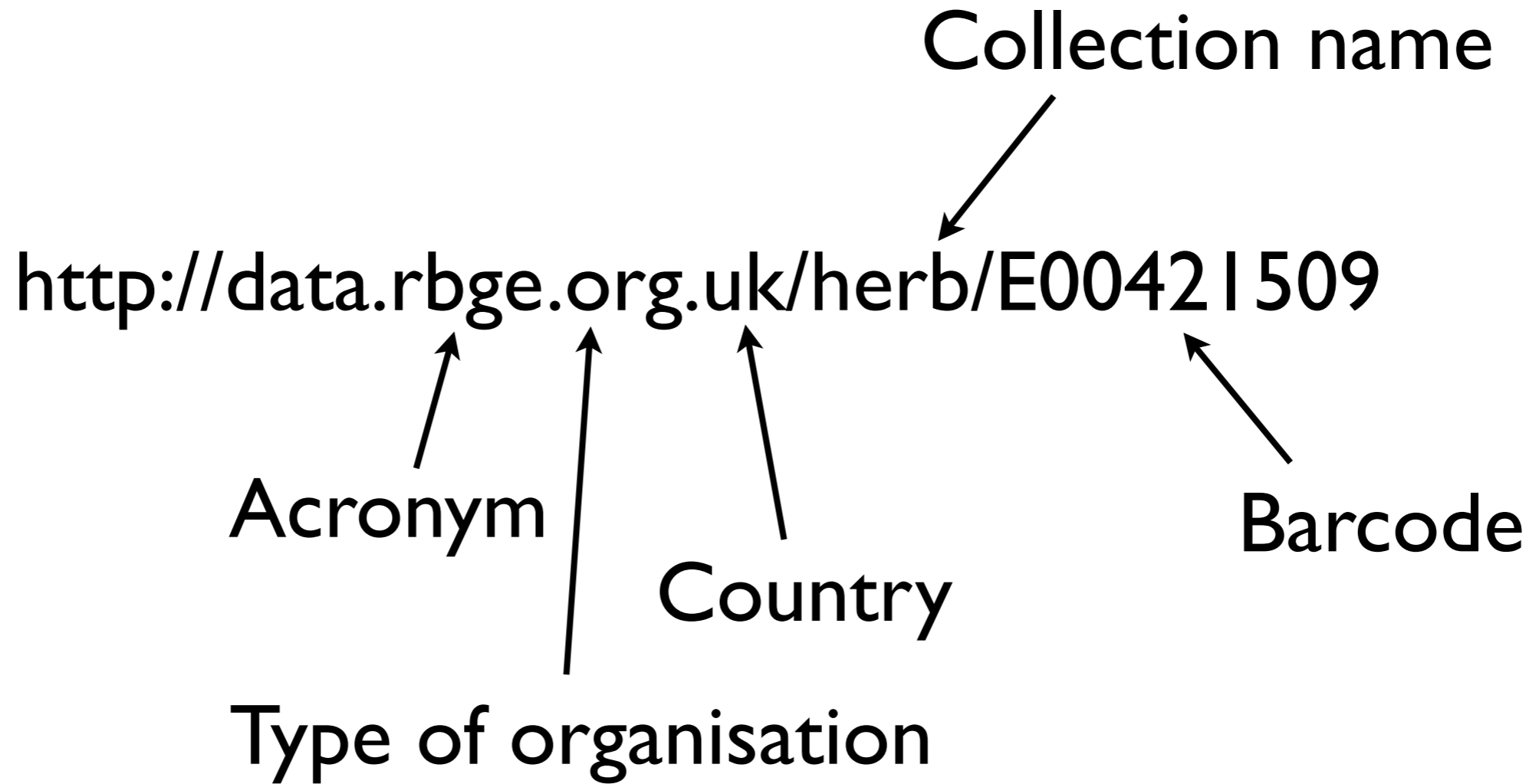
Type of organisation



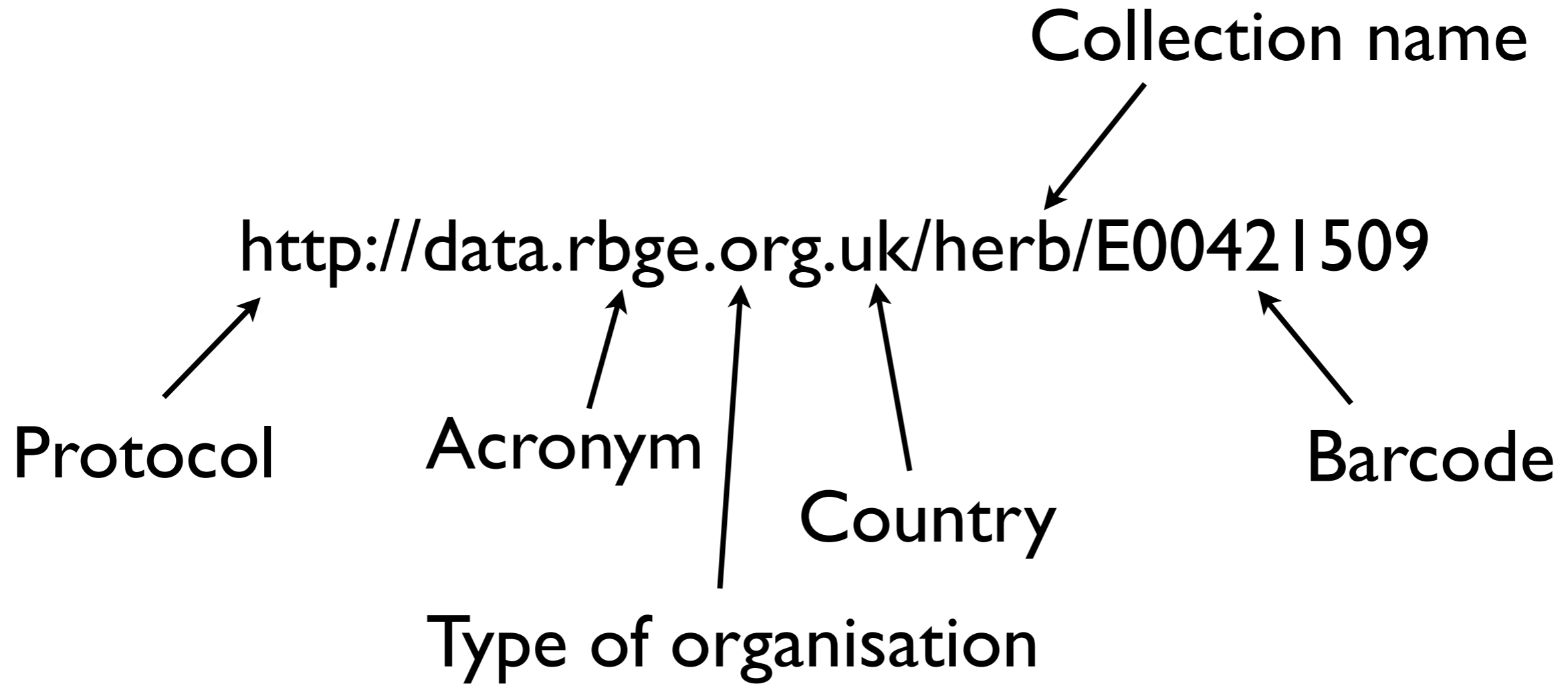
If opacity rocks how come your URIs have attributes in them?



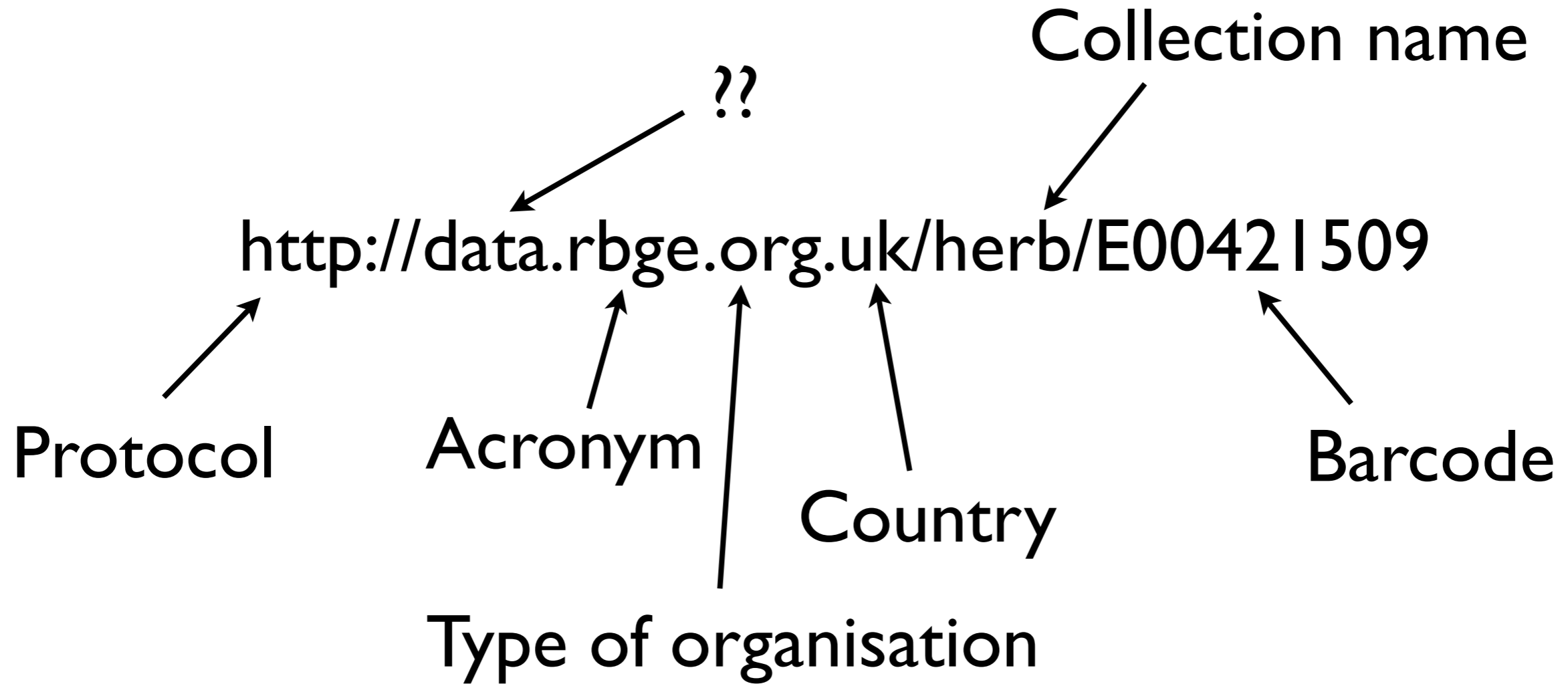
If opacity rocks how come your URIs have attributes in them?



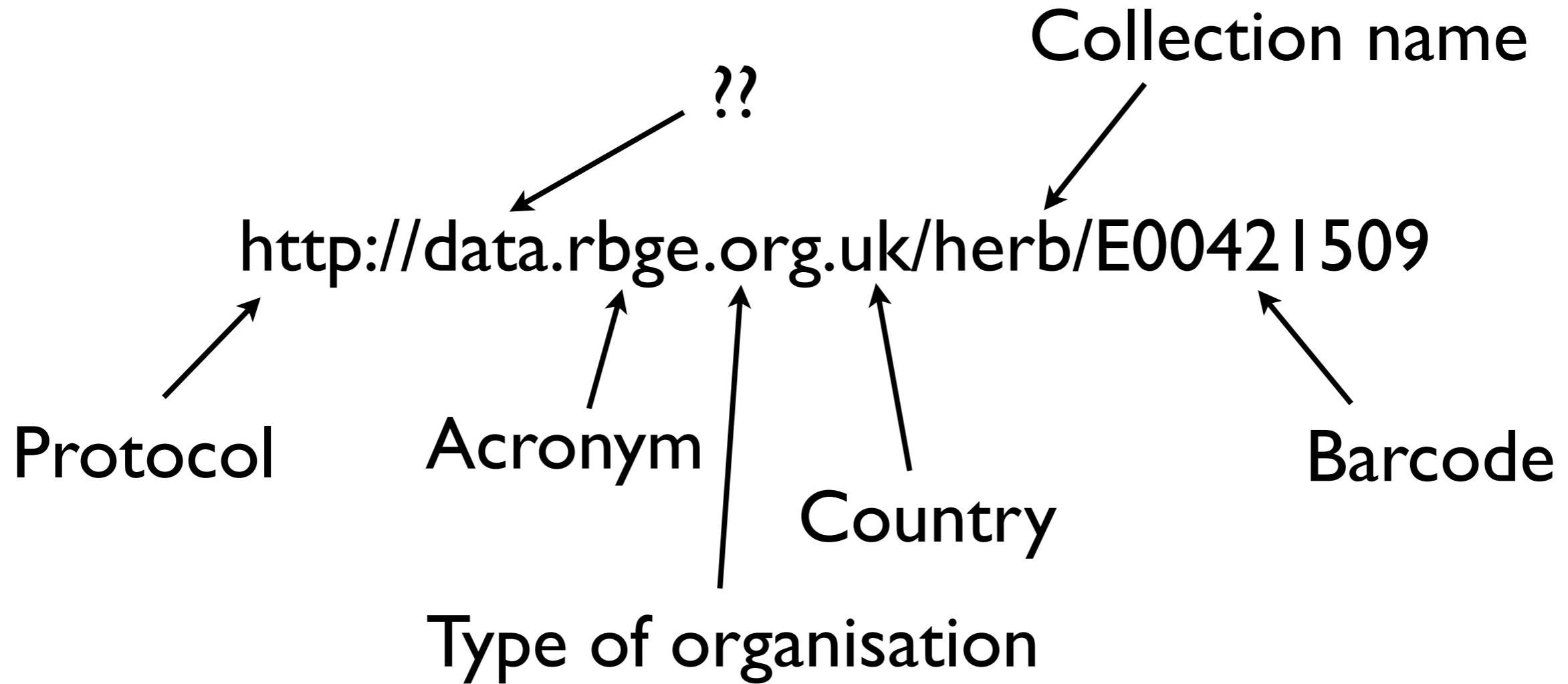
If opacity rocks how come your URIs have attributes in them?



If opacity rocks how come your URIs have attributes in them?



If opacity rocks how come your URIs have attributes in them?



Every part of the string is human readable!

Opacity Sucks

# Opacity Sucks

- <http://9089da05d965.net/f7cf8baa2fd3>

# Opacity Sucks

- <http://9089da05d965.net/f7cf8baa2fd3>
- This would be ideally opaque



# Opacity Sucks

- <http://9089da05d965.net/f7cf8baa2fd3>
- This would be ideally opaque
- Impossible to construct from basic information

# Opacity Sucks

- <http://9089da05d965.net/f7cf8baa2fd3>
- This would be ideally opaque
- Impossible to construct from basic information
- Nearly impossible to type

# Opacity Sucks

- <http://9089da05d965.net/f7cf8baa2fd3>
- This would be ideally opaque
- Impossible to construct from basic information
- Nearly impossible to type
- Difficult to debug

# Opacity Sucks

- <http://9089da05d965.net/f7cf8baa2fd3>
- This would be ideally opaque
- Impossible to construct from basic information
- Nearly impossible to type
- Difficult to debug
- No branding

# Opacity Sucks

- <http://9089da05d965.net/f7cf8baa2fd3>
- This would be ideally opaque
- Impossible to construct from basic information
- Nearly impossible to type
- Difficult to debug
- No branding
- Difficult to get human acceptance

# Middle Way Opacity

<http://data.rbge.org.uk/herb/E00421509>

# Middle Way Opacity

<http://data.rbge.org.uk/herb/E00421509>



Protocol

# Middle Way Opacity

<http://data.rbge.org.uk/herb/E00421509>

Protocol



Existing local scope  
identifier

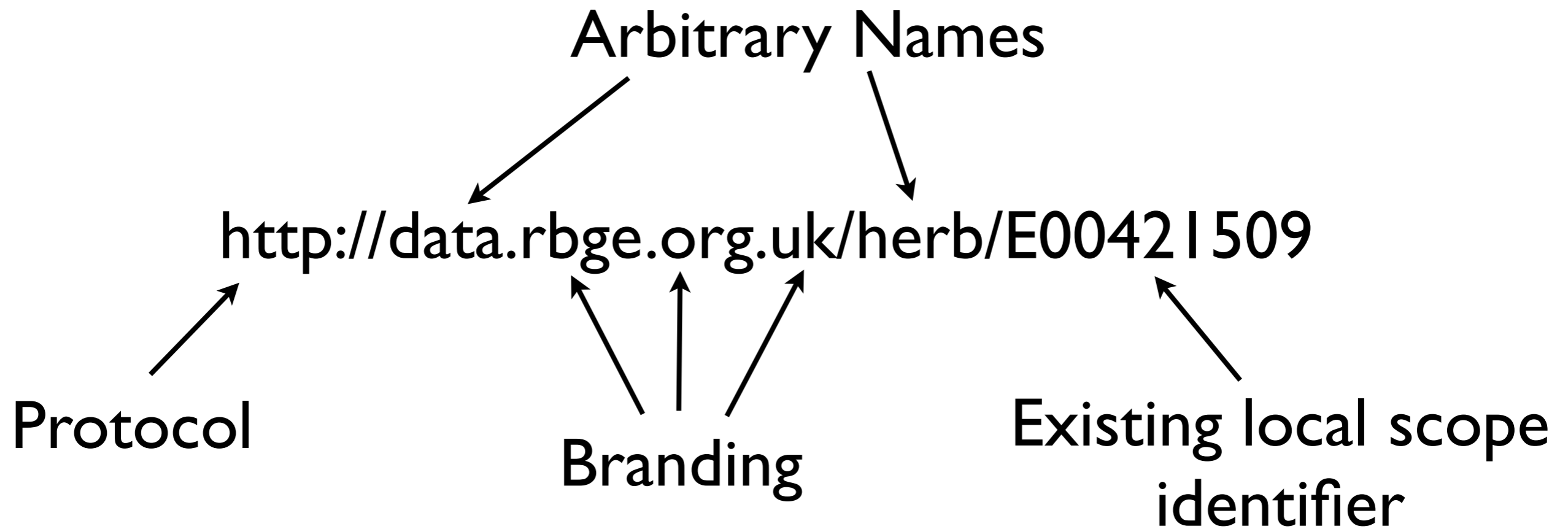




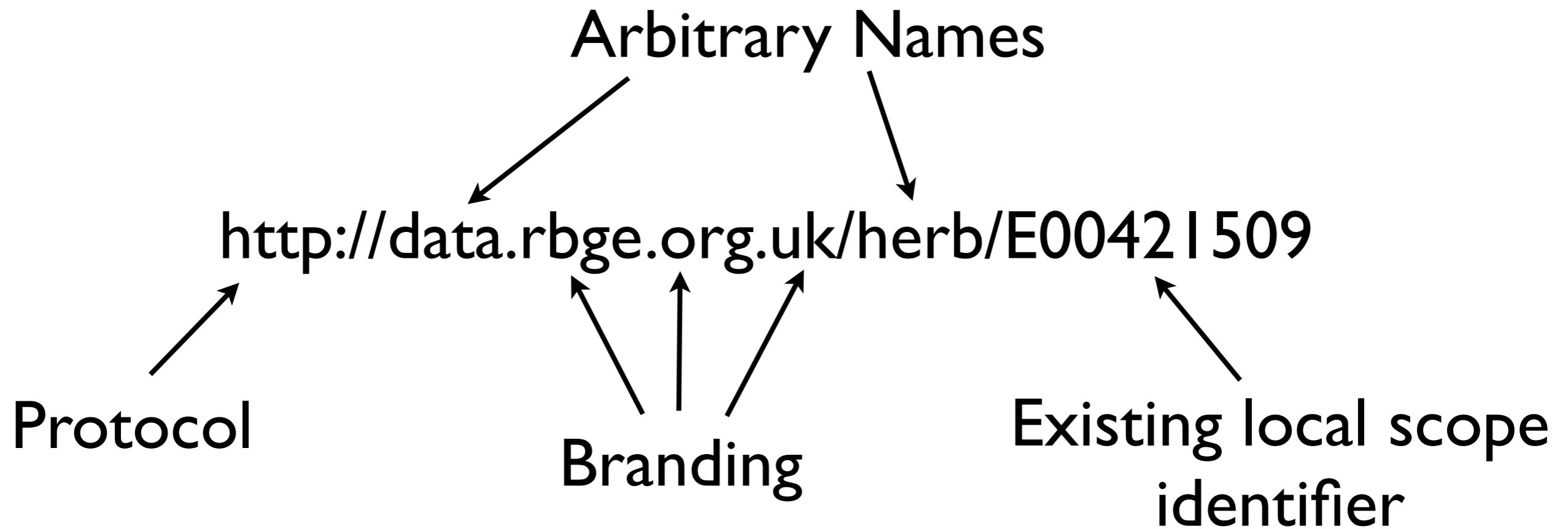
# Middle Way Opacity



# Middle Way Opacity

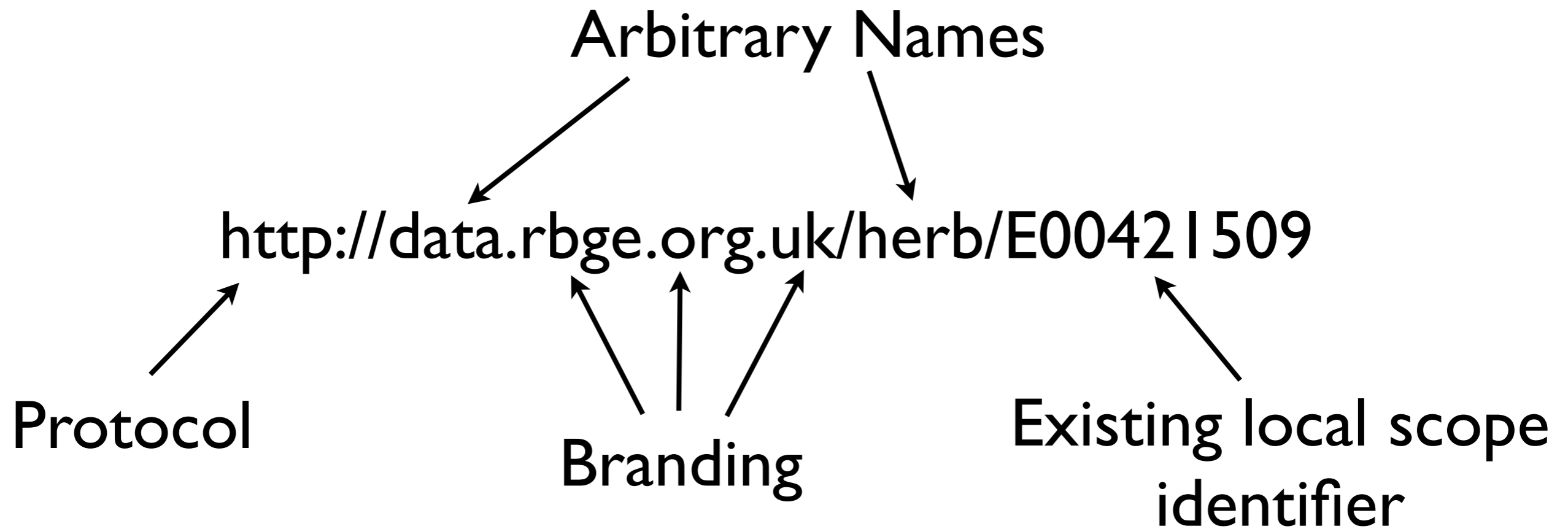


# Middle Way Opacity



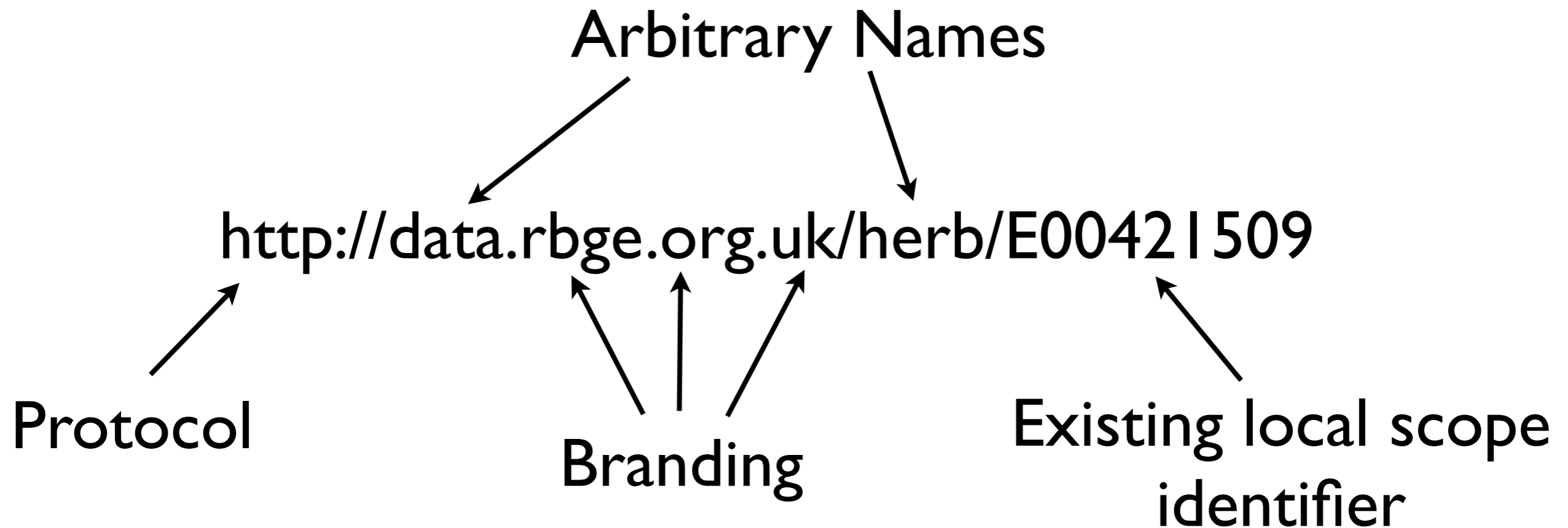
Arbitrary human readable strings

# Middle Way Opacity



Arbitrary human readable strings  
Try to avoid loaded terms

# Middle Way Opacity



Arbitrary human readable strings

Try to avoid loaded terms

Will be immutable!

# Subdomain Balancing

# Subdomain Balancing

- <http://data.rbge.org.uk/herb/E00421509>

# Subdomain Balancing

- <http://data.rbge.org.uk/herb/E00421509>
- <http://herb.data.rbge.org.uk/E00421509>



# Subdomain Balancing

- <http://data.rbge.org.uk/herb/E00421509>
- <http://herb.data.rbge.org.uk/E00421509>
- <http://E00421509.herb.data.rbge.org.uk>

# Subdomain Balancing

- <http://data.rbge.org.uk/herb/E00421509>
- <http://herb.data.rbge.org.uk/E00421509>
- <http://E00421509.herb.data.rbge.org.uk>
- This is a judgement call based on what is best for your institution not a technical decision

# Institution Domain?

# Institution Domain?

- <http://herbarium-e.net/E00421509>

# Institution Domain?

- <http://herbarium-e.net/E00421509>
- Using the institution domain is a good for branding and big collections

# Institution Domain?

- <http://herbarium-e.net/E00421509>
- Using the institution domain is a good for branding and big collections
- Using free standing domain is good for a collection that is not core business of institution and may get shifted (e.g. university department)

# Institution Domain?

- <http://herbarium-e.net/E00421509>
- Using the institution domain is a good for branding and big collections
- Using free standing domain is good for a collection that is not core business of institution and may get shifted (e.g. university department)
- Judgement call not technical decision

# Local Key Choice



# Local Key Choice

- <http://data.rbge.org.uk/herb/E00421509>

# Local Key Choice

- <http://data.rbge.org.uk/herb/E00421509>
- We use barcodes attached to specimens

# Local Key Choice

- <http://data.rbge.org.uk/herb/E00421509>
- We use barcodes attached to specimens
- DB Primary keys probably bad

# Local Key Choice

- <http://data.rbge.org.uk/herb/E00421509>
- We use barcodes attached to specimens
- DB Primary keys probably bad
- Physical attachment to specimen best

# Local Key Choice

- <http://data.rbge.org.uk/herb/E00421509>
- We use barcodes attached to specimens
- DB Primary keys probably bad
- Physical attachment to specimen best
- Should be arbitrary - not collector names & numbers!

# Summary: Balancing Act

# Summary: Balancing Act

- Opaque or Human Readable

# Summary: Balancing Act

- Opaque or Human Readable
  - Only readable enough to manage the URI



# Summary: Balancing Act

- Opaque or Human Readable
  - Only readable enough to manage the URI
- Choosing domains and sub-domains

# Summary: Balancing Act

- Opaque or Human Readable
  - Only readable enough to manage the URI
- Choosing domains and sub-domains
  - Judgement call

# Summary: Balancing Act

- Opaque or Human Readable
  - Only readable enough to manage the URI
- Choosing domains and sub-domains
  - Judgement call
- Choosing object identifiers

# Summary: Balancing Act

- Opaque or Human Readable
  - Only readable enough to manage the URI
- Choosing domains and sub-domains
  - Judgement call
- Choosing object identifiers
  - Just make it arbitrary and easily linked to physical object





Coffee Break