

Linking You Toolkit.....	2
What does this Toolkit look at?.....	2
Why is this important?.....	2
Space-time.....	3
Recommendations.....	6
Shared Vocabulary.....	7
Data.ac.uk.....	8
UK HEI data objects.....	8
Migration.....	8
Further study into institutional websites.....	8
Understand organization challenges.....	8
Complete our study of institutional websites.....	8
URI 101.....	9
Breaking Down URIs.....	9
Best Practises.....	10
Persistent and Permanent.....	10
Consistent.....	10
Readability.....	11
Remove unnecessary keywords.....	11
Query strings should be for filtering and pagination only.....	11
Hashbangs are bad, pushState is good.....	12
Limit the number of sub-domains.....	12
What is a "Cool URI"?.....	13
Why not "URL"?.....	13
Key documents.....	13
Domains.....	14
About www.....	14
Avoid Mystery Subdomains.....	14
Everyone is different.....	15
Easy conformance.....	18
Linked Data?.....	18
Unlock Data's Potential.....	19

# Linking You Toolkit

Linking You was a project at the University of Lincoln funded by the JISC under the 2011 [Infrastructure for Education and Research Programme](#). Its aim was to look at and make recommendations for improving the way that identifiers for .ac.uk domains are planned and managed in higher education institutions.

The Linking You Toolkit is a collection of the outcomes of this project, and is intended as a starting point for institutions looking to better understand and manage their use of identifiers on their .ac.uk website.

## What does this Toolkit look at?

### [Space-time](#)

Mapping the notion of resource identifiers to the physical world of space-time.

### [Recommendations](#)

A summary of our findings and recommendations for excellent identifier design.

### [URI 101](#)

A crash course on what URIs are, how they hold the web together and how they relate to content.

### [Institutional URIs](#)

How URIs on a .ac.uk website relate to specific parts of the institution's operations and structure.

### [Domains](#)

Recommendations for selecting and maintaining well structured domains and subdomains.

### [Institutional URIs](#)

Our source data and findings on how institutions currently arrange their identifiers.

### [Data Model](#)

Our recommended URI structure, with a look at how it fits into Linked Data.

## Why is this important?

The web is now fundamental to the activity and idea of the university. This Toolkit provides a standard way of thinking about your institutional URI structure, making it easier for people (and their browsers) to both remember your web addresses and locate where they are in your web site. It also helps prepare your institution for the world of linked data by proposing a clear and concise model for your data, making smooth integration with other systems easier and faster. A good URI structure can be easily understood by both humans and machines.

This may seem like an obscure technical project to many people, but we think there will be a number of benefits, both internal and external, in having undertaken this work. We've situated this project within the wider context of the [Learning Landscapes](#) project at Lincoln and have tried to think of identifiers as expressing and sign-posting new forms of spatiality. Just as traditional space architects offer easy to understand blueprints so that people can imagine the space, Linking You has offered a blueprint for institutional identifiers. We don't expect every institution to adopt the exact same syntax, but we do hope that it leads to a new shared vocabulary for institutional identifiers. The visual representations created during the project will be used at Lincoln to communicate proposed changes to the online research, teaching and learning environment. Furthermore, our work on the project will influence our provision of open data allowing for greater re-use and personalisation of information for students and staff. We hope the Toolkit provides similar benefits to you, too.

Although URIs are [increasingly being obfuscated](#) by developments in browser design, they are also increasingly being integrated into browser search functionality such that the benefits of providing clear, plain language URIs has never been greater. As well as providing subtle benefits to your website visitors (and your SEO), technical staff at your university should benefit from a consistent approach to URIs so that it is significantly easier for them to manage the relationships between resources, as well as making it simpler to produce documentation which refers to URIs that *make sense*. Alongside this, your University will be able to implement an efficient method of redirecting users to their intended destination when a resource has moved.

Institutions as a whole will benefit from thinking about a logical and human readable URI addressing system for online services. If implemented, they should experience:

- Better IT systems integration.
- Improved navigation of virtual spaces.
- Appropriate conventions for differing technology platforms i.e. mobile/desktop devices.
- Future proofing against non-sustainable URI management practices.
- Ability to 'design with data'
- Improve discoverability of resources (and SEO)

We hope that you'll benefit from our project documentation, the comparative work we've done and our data modelling. Linking You was a small project that we hope will offer useful contributions to a wider conversation across the sector.

## Space-time

"What, design a URI? I have to design URIs? Yes, you have to think about it."  
*Berners-Lee, [Cool URIs don't change](#), 1998*

During 2007-9, the University of Lincoln led the HEFCE-funded [Learning Landscapes project](#) which offered "the higher education community a practical

and conceptual framework to consider the ways in which learning and teaching spaces are being designed and developed." The Learning Landscape project provided 12 case studies of different UK universities that had recently designed and built new learning spaces, as well as a set of development tools to support the inclusive work of designing effective learning spaces between academics, support services, other key stakeholders and students. The project acknowledged the essential role of technology in the design of the learning landscape, as well as recognising the extent to which the Internet has disrupted and extended our understanding of what constitutes a learning space. The remit of the Learning Landscapes project was very much focused on the communicative process between different stakeholders around the construction of physical space and the extent to which technology might alter and enhance our understanding of spatiality.

With the Linking You project, we've attempted to draw from the recommendations of our earlier project and consider in much more detail one aspect of what might be called the 'virtual learning landscape' but we would argue is as *real* and as *valuable* to the work of Higher Education Institutions as the physical real estate. In our project, we understand technology not as an enhancement to the learning landscape but intrinsic to it ([Feenberg 1999](#)). This has always been the case ([Selwyn 2011](#)) but with the development of the Internet, we've tended to understand this particular set of technologies as something apart from the physical world, something 'virtual' and therefore somehow secondary - an after thought.

The overall output of the Linking You project is this toolkit, a plugin to the original Learning Landscapes project tools, which is meant to both support and facilitate a better understanding of institutional URIs and act as a focal point for both technical and non-technical stakeholders. We want you to trash your notions of the 'virtual', *get real* with URIs and understand their intrinsic role in developing and managing a contemporary university. We hope this is a toolkit that will help you reverse engineer **.ac.uk** and *imagineer* your domain in terms of the very *idea of the university*.

When talking about something as complex as 'the university' we're assuming that we understand what a university is *for*. The Learning Landscapes project tools have been useful to us in providing context to our specific task of examining the use and value of URIs in the university sector. The first tool challenges us to re-imagine higher education through an understanding of the history of the idea of the university. In other words, *know your domain*. Looking backwards at the idea of the university is referred to as a process of 'reverse imagineering', an idea influenced by the practice of 'reverse engineering' software, and this is something we have attempted to do through the comparative study of 40 institutions' and then the development of a proposed model of best practice.

In the current climate, as universities face difficult financial and ideological challenges, the idea and ideal of the university is being increasingly questioned. Today's university retains traces of its long history and this is no less apparent in the way that HEI's organise their resources and construct their URIs. The

distinction first made 200 years ago between teaching and research is retained today and despite [efforts](#) to integrate these two core activities of the university, institutional websites clearly separate the two and more often than not reflect the increasing merger of research and enterprise. This is no surprise in the age of the 'entrepreneurial university', but would it be so strange for our institutional URIs to act as sign-posts across another space-time dimension of the campus, with the potential to extend our traditional collegiality to the so-called 'edgeless university'? ([DEMOS 2009](#)) Taking this point of view, the mirroring of the physical structure of the university onto the web architecture of the institution can be understood as a rational but limited approach to the use of the web, reinforcing the idea that web development is simply an enhancement activity rather than an opportunity to re-architect university space-time.

“And now the campus is massive, one of those dominant modern environments of multi-functionality that modern man creates: close it down as a university...and you could open it again as a factory, a prison, a shopping precinct.” *Malcolm Bradbury, [The History Man](#), 1975, 69*

Is it any surprise that institutional URI schemes so closely reflect the idea of the university that has developed since the medieval period? If, as Malcolm Bradbury wrote in the 1970s, the modern campus is so lacking distinction that it could just as easily serve as a prison, factory or shopping precinct, are our institutional websites condemned to the same indeterminate features? If the idea of the university is in danger of being lost, can it be re-asserted through the new spatiality of the Internet, which increasingly feels like an "annihilation of space by time"? ([Marx, Grundrisse Ch. 10](#)) The economy of *land*, so ingrained in our conception of the world, seems barely relevant to the architecture of the web, but reveals itself continually in debates around Intellectual Property and copyright. No wonder that the avant-garde of web architecture is the 'open web', the 'web of data', whose proponents in just the past couple of years have increasingly argued for public domain licensing (e.g. [CC0](#), [PDDL](#)) and the effective abandonment of property as the organising principle of the web.

“The logic of digital technology leads us in a new direction. Objects, as well as ideas, are no longer fixed, no longer tangible. In cyberspace, there is no weight, no dimensions; structure is dynamic and changing; size is both infinite and immaterial. In this space, stories are written that change with each new reader; new material can be added, and old material can be deleted. Nothing is permanent.” *Kleinman, [Don't Fence Me In: Copyright, Property and Technology](#), 1996, 76*

This new space - cyberspace - allows for a new science of space, a *spatiology* that allows us to think critically and imaginatively about the idea and form of university we desire. We discussed this in the Learning Landscapes project, highlighting how critical pedagogy can be used as a design principle, a resource in the design and construction of a counter-space, providing critical tools with which we reverse engineer the university. Cyberspace allows for 'Utopian thinking' through which the constraints of property rights, and the traditional hierarchies of research, teaching and learning become "manifest as entirely

different spatial forms and temporal rhythms." (Harvey, [Space of Hope](#), 2000, 237-8) Arguably, we're already seeing this Utopian thinking in the forms of the Open Data, Open Access and Open Education movements. This was recognised two years ago by JISC in the [Edgeless University](#) report and it's time that the foresight of these critical movements shifts into the mainstream and their Utopianism infects the whole of .ac.uk. How do we begin to do it? Read our [recommendations](#).

## Recommendations

In this toolkit, we have argued for clear, consistent, readable and accessible URIs: *Cool URIs*. Cool URIs don't necessarily assume that the resource being represented is fixed, stable or permanent, only that the URI persists, identifying a resource (or more likely a collection of resources) that we acknowledge may change [like the weather](#).

By proposing our ideal model, we are not expecting each institution to redesign the content of their websites, but rather we have highlighted where the institutions in our study could redesign and remap an improved underlying URI structure. **This is an important distinction:** the design of an institutional URI structure has little to do with the resources as are they are presented to users in their web browsers. **Website design is not the issue here. URI design is.** Although we are [calling](#) for a much more imaginative, experimental and ambitious use of the .ac.uk domain, if planned well, work on institutional URI structures could take place without anyone actually noticing thanks to [HTTP 3xx status codes](#). This alone would improve the opportunities for data aggregation and re-use and the development of services built on simple APIs.

An abstract space that is perfectly organised relies on the agreement and co-ordination of a number of technologies (e.g. networking, web servers, browsers). In our physical spaces, technologies work together so as to ensure the appearance and experience of order and continuity. The same applies to the design and development of our cyber spaces when there is an appreciation of how each technology can conceal the complexity of the other. For example, web servers can hide file extensions and seamlessly map one resource onto another, URIs can be constructed according to a grammar that is meaningful over time, browsers can integrate search features into the location bar and so on. The abstract space of the URI should be thought about and designed so that it is not only perfectly organised but persistent (trustworthy, reliable) and offers us a constructive grammar of signs which can be understood by human and machine.

The Learning Landscapes project asserted that the three fundamental qualities of good design are efficiency, effectiveness and expression. These are communicative attributes, describing an ideal space that is shaped by a number of recommended principles that are broadly applicable to cyberspace, too. When thinking of university cyberspace for the Linking You project, those principles on page 46-7 of the [Learning Landscapes Final Report](#) may be reformulated as:

- Drive research into the effective design and development of university websites
- Provide support to teachers and students for Utopian thinking and experimentation on the web
- Include students as clients and collaborators in the design of university websites
- Be academically credible. Web development should not simply be a technical exercise removed from the academic rigour of the university
- Understand the relationship between space and time: it's not just 'cyberspace', but space-time
- Articulate the institution's vision and mission as a connected, networked whole
- Create incentives. Recognise and reward innovation across all staff and students
- Create formal and informal management structures that support strategic experimentation and imagineering (e.g. 'think tanks', 'sand pits', 'skunk works')
- Avoid stereotyping. Bring people together from across subject areas and professions so as to avoid an 'us and them' attitude
- Intellectualise the issues. Generate debate on the nature of academic values and the role and purpose of higher education: the idea of the university is synonymous with the idea of .ac.uk

In essence, much of this could be broadly understood as 'digital literacy', where the learning ~~land~~cyberscape is designed to engender capable, confident and critical individuals engaged in research, teaching and learning, so that they are [active producers](#) of their own social world. It feels like the work required to really understand and take advantage of our new university space-time has hardly begun.

In addition to these principles, the Linking You project has its own recommendations, too.

### Shared Vocabulary

The first is the need for a shared vocabulary - one that is understood and used by academic and managerial stakeholders within universities and across the sector. One of the key aims of the Learning Landscapes project was to "provide a clearly understood vocabulary within which the future development of the University can be articulated, in order to better inform the design of the built environment of higher education." This same approach is common across many other domains, not least that of web architecture, and our comparison of 40 institutional websites has identified where there is already agreement around a shared vocabulary and where there is divergence. In short, we'd like to see the data model in our toolkit be developed into an ontology for the HEI sector. We've made a start with our model, but we'd like to see a national working group that includes academics, professional services and students.

## Data.ac.uk

Second, we'd like to see this proposed vocabulary inform and be informed by the work being done on data.ac.uk, ideally adopting the ontology as the definitive model for organising data.ac.uk. This would not only demonstrate that the ontology has formal backing but also encourage contributors to data.ac.uk to use the ontology in their own institutions.

## UK HEI data objects

Third, we think there's mileage in scoping out and prototyping an open directory of UK HEI data objects e.g. courses, people, research centres, locations, etc. This could act as a halfway house for institutional web managers not able or willing to contribute to data.ac.uk but do wish to publish their URI endpoints for the convenience of developers and researchers. Our comparative study of 40 institutions is a start (and we intend to add to it), but something more sustainable needs to be put in place.

## Migration

Fourth, attention needs to be given to the way institutions transition to a shared ontology for the sector. Research needs to be done that examines and recommends strategies for migrating from existing and legacy URI structures to a model of best practice. HTTP 3xx status codes are at the heart of this.

## Further study into institutional websites

Fifth, we think that the work of the Linking You project could be extended to look at the use of other areas of institutional websites. While we concentrated on our 'corporate' website, we recognise that this is just one aspect of university cyberspace and more needs to be done to understand what makes up the whole. This could be tied in with a study of institutional use of CMS products, which would give us a better appreciation of the technical barriers and opportunities for change that are available to HEIs.

## Understand organization challenges

Sixth, to complement the technical focus of our recommendations, we need to better understand the organisational challenges involved, too. We therefore recommend that work is undertaken to survey, interview and case study web managers' and other stakeholder attitudes around the use and 'value of URIs'.

## Complete our study of institutional websites



Finally, a couple of extensions we've given ourselves on this project: One is to further our study of 40 institutions to complete the national picture of over 100 HEI websites. We began the project by thinking we'd compare our own URI structures with a couple of other institutions, then we increased this to looking at 20. When we hit 20, we realised that 40 wouldn't be too difficult to do and might give us more confidence in our proposed model. Now, having hit the deadline of the Linking You project, we're still keen to do more and intend to complete a national comparative study before the end of the academic year. Secondly, we also intend to reformat this toolkit as a Briefing Paper and send one to stakeholders in every institution along with a poster of the proposed model to stick on their walls.

## **URI 101**

An obvious place to start looking at institutional identifiers is by looking at what URIs actually are, how they work, and what the current best practice is when it comes to designing them.

URI stands for Uniform Resource Identifier, a set of characters which uniquely identify a specific resource on the internet. URIs consist of two parts, a scheme which defines how a resource can be contacted, followed by a colon and then a scheme-specific part which uniquely identifies the resource. An example of a URI might be `http://google.com`, where `http` is the scheme and `google.com` is the identifier. There are many different URI schemes available which transfer a specific type of content, or transfer it in a specific way. A few examples of common schemes are HTTP, HTTPS, FTP and Mailto, although there are a great many more available.

Linking You focusses specifically on the HTTP (and by extension HTTPS) scheme, since this is the default scheme over which web content is transferred. However, a great many of the suggestions in this toolkit will apply to other schemes.

## **Breaking Down URIs**

A HTTP URI is made of two parts, one of which (generally) identifies the server and the other of which (generally) identifies a resource on that server. Exceptions to this will always exist due to the fluid nature of the web — for example where an institution uses complex load balancing — but the assumption will apply to the vast majority of cases. An example URI would be `http://example.com/section/resource.htm`. In this URI we can begin by looking at the scheme, in this case HTTP. This leaves `example.com/section/resource.htm` as the unique part of the identifier; the part which identifies a resource. Everything up until the first slash is the server address or domain, in this case `example.com`. Everything after the slash identifies the resource within that domain, in this instance `section/resource.htm`.

In the case of academic institutions the domain is mostly unchangeable, being inextricably linked with the institution to which it is associated. A change of domain name for an established institution would be extremely rare, and only happen should the institution substantially change its name (such as the University of Lincolnshire and Humberside becoming the University of Lincoln, and ulh.ac.uk becoming lincoln.ac.uk). However, the resource identifier is easily altered by changing the configuration of a web server. This toolkit therefore focusses on the resource identifier part of the URI, with the exception of the [Domains](#) section.

## Best Practises

Developing a website, small or large, isn't a small undertaking. Below are a number of best practises to bear in mind when it comes to developing the URI structure.

### Persistent and Permanent

One of the most fundamental philosophies behind a URI is that it represents a data object on the Internet. The URI must be unique so that it is a one-to-one match - one URI per one data object.

While this is always the goal, there are times at which it is very difficult or impossible to accomplish. Canonical URL tags were invented to help reduce the amount of duplicate content seen by a search engine. While not a final solution, canonical URLs are strongly recommended as large search engines like Google are now paying attention to them.

URIs should also be permanent (i.e. choose the URI once and leave it at that). This speaks to good URI design before a site is launched, with the URIs being carefully planned. There will come a time when you do want to make improvements to your choices or otherwise must change URI structure. When this becomes a necessity, HTTP 301 moved permanently redirects should be set up. This tells browsers and search engines the new location of the content and will also preserve any Google PageRank (and other search engine rankings) that the old URI has accumulated.

### Consistent

URIs across a site must be consistent in format. Once you pick your URI structure, be consistent and follow it! Having good URI structure for part of the site means that you still have poor structure overall. In order for a user to trust that URIs work a certain way on a site, the format must be consistent. If you must switch structure (maybe you're updating a poorly-designed site), use 301 redirects as previously mentioned.

An example of poor consistency would be having undergraduate information located at /undergraduate and postgraduate information at /study/postgraduate.

Consistent structure = maintainable websites.

## Readability

A URI can simply be used as a 'click to' point on the internet. There's nothing stopping an HEI putting a page on courses in the School of Computing at <http://example.ac.uk/bcwi83b>. You plug it into a link, people click the link and off you go. Technically this is sound, but only in the same sense that you can technically address a letter to something like "10, SW1A 2AA". Yes it's compact and yes it works, but it conveys absolutely nothing in terms of context. It's also a real pain to remember, and requires you to use additional bits of your brain if you're ever writing it down for later reference or typing it into a browser address bar.

Imagine for a second that a prospectus had the following:

Find out more about Computing at <http://example.ac.uk/bcwi83b>

And then compare it with a 'human' address:

Find out more about Computing at <http://example.ac.uk/school/computing>

Now, try to remember the first one without looking at it. In short, URIs should describe your content (but in a short and succinct manner).

Whatever method you use to create your website, it must be able to generate human readable URIs. Increasingly web browsers are allowing people to [search through their history](#) based on URI fragments, meaning that while a URI such as [/computing](#) will be easily found, [/\\_depts/cs](#) won't be. Even worse would be the style of URI which is often created by an incorrectly configured CMS, such as [ac.uk/content/027463](#).

## Remove unnecessary keywords

Following on from the above point about readability, it is possible to take clean URIs too far by including unnecessary information in the URI.

For example, an about page with the URI [/about\\_the\\_university](#) whilst accurate and readable, could and should be shortened to [/about](#) because it doesn't lose any of its meaning but still describes the content well.

Watch out for content management systems generating URIs based on page titles. A URI should be made up of keywords, but only the absolute minimum number of keywords in order to describe the content.

## Query strings should be for filtering and pagination only

"Dynamic" URIs, i.e. URIs containing query string arguments such as `/content/page.php?id=1234&output=1` should be kept to an absolute minimum, and even then should only be used (if necessary) for filtering content and paginating.

Dynamic URIs are less readable to both humans and search engines and therefore could be seen as less trustworthy because they don't necessarily describe the content to the user.

If you need to use query strings then try and ensure they are descriptive:

e.g. `/undergraduate/courses?years=2` or `/events?page=2`

### Hashbangs are bad, pushState is good

A number of websites, including Twitter and Gawker Media (Lifehacker, Gizmodo, etc) have recently re-architected their websites to make use of hashbang URIs — e.g. `https://twitter.com/#!/unilincoln`. [The hashbang was recommended by Google](#) as a way for search engines to crawl AJAX powered websites.

There are a number of problems with hashbangs:

- In order to decide which content to render based on a hashbang URI, a hashbang enabled website relies on a user having a modern JavaScript enabled web browser.
- Hashbangs are invisible to the server, so if someone visits `http://example.ac.uk/#!/badurl` (which triggers a 404), the error will not appear in your server logs.
- Hashbangs are forever, so if you go hashbang you can't go back. You can control the links on your website, however you can't control other people's links to your website. If people start linking to your hashbanged URIs then you're going to have to support the parsing of hashbangs even if you implement "fixed" URIs again.

If you want to be all modern and exciting you should make use of the new [pushState](#) JavaScript features being introduced in the latest versions of browsers.

For example, if you you want to move from `http://example.ac.uk/undergraduate` to `http://example.com/undergraduate/courses` you'd provide a link, which when clicked, would AJAX load in the content from the other page and update the URI in the navigation bar, or for a user with an older browser it would just load the other page as normal.

### Limit the number of sub-domains

There may occasionally be a valid reason that content is on a different subdomain compared to the rest of the content, and with appropriate linking

between the two should be okay. However if you split all your content subjects across lots of subdomains e.g. home.example.ac.uk (as opposed to example.ac.uk/home) and there aren't all of the correct 301 redirects in place then users will get confused and frustrated, and search engines will just give up on your site.

There is a very well known principle called KISS (Keep It Short and Simple) that is very appropriate in this situation.

## **What is a "Cool URI"?**

This toolkit is recommending something called "Cool URIs". This is an important part of creating understandable, reusable and memorable addresses.

A cool URI is one which is entirely human readable, and stripped of as much 'technical' information as possible. Where an 'uncool' URI might be something like `/_course/J933.aspx` a 'cool' URI would be `/course/audioproduction`. Although the length may be similar, the cool URI is much more readable and helps to orient the user in your web space.

Uncool URIs are commonly found on sites using legacy or badly configured CMS systems, where a page may be represented by a seemingly random identifier such as `/index.asp?content=64927`. There are many technical methods for dealing with uncool URIs and transitioning to the recommended cool alternatives. Linking You recommends the use of HTTP 3xx status codes — such as HTTP 301 Moved Permanently — as these deal seamlessly with users using the old URI, as well as keeping search engine rankings mostly unaffected.

## **Why not "URL"?**

You may be more familiar with referring to URIs as URLs. This isn't (for the most part) wrong and the distinction is mostly academic, but the generally accepted term in technical publications by Internet standards bodies is to use "URI".

For a more in-depth look at why we're using URI instead of URL we recommend taking a look at [RFC 3305](#), a document jointly published by the W3C and IETF standards groups.

## **Key documents**

[Hypertext Transfer Protocol -- HTTP/1.1](#)

[Cool URIs for the Semantic Web](#)

[Cool URIs don't change](#)

[Uniform Resource Identifier \(URI\): Generic Syntax](#)

[The Self-Describing Web](#)

[Architecture of the World Wide Web, Volume One](#)

## Domains

Domains are a key part of the Internet, allowing people to enter a single, easily memorable string of characters (such as google.com) into their browser to reach a resource, rather than a seemingly arcane string of numbers. Domains in higher education tend to end with the .ac.uk. characters, a subset of the domain space which is managed by JANET. Each institution will generally have one root domain, such as lincoln.ac.uk, within which they can create and manage their own subdomains to allow people to connect to individual services.

Whilst *Linking You* focusses mostly on the identifiers which are part of a domain (predominantly within the root of the domain), it's worth looking at a few best practices and suggestions for keeping the domain space clean and easily understood.

## About www.

Everybody is used to seeing the letters www. at the start of a web address, but they're not actually a requirement. Their presence is historical, stemming from when it was necessary to prompt users to enter the domain in their web browser, implying that the website was part of the "World Wide Web". Since users are now much more familiar with the process of visiting websites, any use of the www. prefix can be considered redundant.

*Linking You* recommends that websites at least support accessing them without the www. prefix. You can find arguments for this approach, as well as necessary code snippets, at the [no-www campaign](#). The transition to supporting a domain both with and without the prefix is technically very simple, and does not require any significant amount of work.

Making the www. optional has many benefits, including faster reading of the web address (it has ten fewer syllables), requiring less space on printed or advertising material, removing redundancy and making the address easier to type (in the case of short root domains such as ox.ac.uk adding the www. makes the address a whole 50% longer).

## Avoid Mystery Subdomains

It may be tempting to create subdomains with long-winded descriptions, internal acronyms, codenames or similar. This should be avoided wherever possible. If you've got a service which allows people to top up their printing credit try to give it a domain such as print. or topup., rather than topupmyprintaccount. or onlinemoneyloader. The shorter domains are far easier to remember, and much more intuitive.

This data was collected during the final days of May 2011 — institutions may have modified their URI structures since this point.

The data represents key areas of a total of 40 higher education institutional websites, as well as highlighting the differences that institutions have in URI structure. We studied 10 institutions from four of the UK university collaboration groups — the 1994 Group, the Russell Group, the Alliance Group and the Million+ group.

The red highlights represent where a website (at the time of the survey) goes against Linking You recommended practices, and where such a behaviour is heavily discouraged for technical or usability reasons. In the case of the "WWW Required" column it means that the site is not accessible without the www. prefix (see [Domains](#) for more on this), and in the case of the "404 Redirect" column it means that the behaviour effectively breaks the Internet's expected behaviour model.

Where a column has green highlights they indicate that the URI for that area follows the recommended Linking You "Cool URI". However, where a column is not highlighted it does not indicate a problem with the URI. In many cases different URIs exist for historical or local reasons and - while not what we'd recommend for maximum cross-institution compatibility - are entirely valid.

## Everyone is different

As this is the first study of its kind into HEI website URI structures, and there are no pre-existing recommendations, we were expecting little correlation between the websites however for in some cases this isn't the case.

We discovered there is very little consistency about the URI structure for undergraduate and postgraduate information and course pages, and especially for postgraduate course information there is no agreement whatsoever, with only one institution - Kingston - with a URI that follows the structure we're recommending (/postgraduate/courses).

However, for some of the more "static" pages such as research and business services, most institutions seem to have followed best practise (and again, what we are recommending in our model) by using /business and /research as the namespaces. Among those institutions using alternative URIs for these pages there appears to be a mixture of content management systems influencing the URIs (e.g. Reading: /working-with-business) and alternative terms such as /enterprise and /partnerships.

One interesting trend can be found in the about pages. Most institutions go with the URI /about, however a lot of universities have their abbreviated name in the URI, e.g. Nottingham Trent: /about\_ntu, and De Montford: /about\_dmu. Again, this could be down to the content management system rendering URIs based on page titles, but in any case this is an example of having unnecessary information in the URI which goes against best practice.

Also a number of institutions place their contact page inside the /about namespace which seems illogical because contacting the university is a call to action and therefore deserves its own top level URI at /contact.

We also examined some other important aspects of websites such as presence of a sitemap and error redirection and the requirement of a www. prefix to the domain.

Sitemaps allow both humans and search engines to discover content across your site and they should always be in the root. A number of websites we examined do have sitemaps however they call them “a to z” or indexes which are accurate but really should follow the correct web terminology which is “sitemap”.

Error redirections are a very important issue. If I went to /foobarfooey, which is unlikely to be a real web page, then I expect the web server to issue a 404 webpage with a 404 header. Likewise if there is a server error (error 500) or I visit a web page I’m not authorised to access (error 401), the server should keep me at the address I visited (in this case /foobarfooey) and not redirect me to a dedicated error page. Websites that redirect on error break the web because they use a 301 (moved permanently) or 307 (temporarily moved) header which means that the content existed at some point (which is untrue) and the content can now be found at the dedicated error page which in-turn issues a 404 content not found header. This is really bad practice.

In the toolkit we have discussed that www. prefixes are unnecessary in this day and age. In our study we found most sites we looked at worked well without the www. prefix to the domain name however a quarter of those sites fail completely if the www. prefix is missed off. The remaining sites force the use of www. (or another subdomain such as www1., www2., or home.) which at least is forgiving. We recommend that all institutions update their domain DNS settings to support no prefix. This is as simple as creating a CNAME to point at the root domain: e.g. CNAME www.example.ac.uk. → example.ac.uk.

This is the HEI URI model we are recommending to the sector as a result of our analysis of existing institutional website structures and also our own informed opinions on URI structure.

As developers we are used to the concept of objects and key/values and this is reflected in the model's structure. We've gone for namespaced type/identifier object nodes, such as /event/openday/{identifier}, /about/campuses/{identifier}, and /academic\_departments/{identifier}/courses.

The only contention we had on the model was whether or not to have courses as a top level object, or have them as a sub-object of student type (undergraduate/postgraduate/foundation). In the end we went with sub-objects as this seems to be a standard that other HEIs have agreed on because postgraduate courses are structured and administered in a different way to undergraduate courses. Despite this, individual courses should still always be



forwarded to the `/course/{id}` resource, as a course in itself has no undergraduate or postgraduate status outside of the institution's own administrative structure.

`institution.ac.uk`

- `{ucacode}` → Redirect to appropriate `/course/{id}`.
- `/courses`
- `/course/{id}`
- `/undergraduate`
  - `/courses`
    - `{id}` → Redirect to appropriate `/course/{id}`.
    - `/search/{query}`
    - `/entry_requirements`
  - `/prospectus`
- `/postgraduate`
  - `/courses`
    - `{id}` → Redirect to appropriate `/course/{id}`.
    - `/search/{query}`
    - `/entry_requirements`
  - `/prospectus`
- `/foundation`
  - `/courses`
    - `{id}` → Redirect to appropriate `/course/{id}`.
    - `/entry_requirements`
  - `/prospectus`
- `/events`
  - `/opendays`
  - `/conferences`
  - `/public_lectures`
  - `/graduation`
- `/business`
  - `/incubation`
  - `/ktp`
- `/research`
- `/academic_depts`
  - `{id}`
    - `/courses`
      - `/undergraduate` → Redirect to `/undergraduate/courses/search/{query}`.
      - `/postgraduate` → Redirect to `/postgraduate/courses/search/{query}`.
  - `/staff` → Redirect to `/contact/search/{query}`.
  - `/news` → Redirect to `/news/search/{query}`.
- `/support_depts`
  - `{id}`
  - `/staff` → Redirect to `/contact/search/{query}`.
  - `/news` → Redirect to `/news/search/{query}`.
- `/about`

- /vc
  - /parents
  - /{city}
  - /campuses
    - /{id}
- /search
- /press
  - /facts
- /news
  - /{date}
  - /{id}
  - /search/{query}
- /jobs
- /legal
  - /policies
  - /data\_protection
  - /environment
  - /equality
  - /foi
  - /ict
  - /website
  - /regulations
- /contact
  - /staff
  - /{id}
  - /search/{query}

## Easy conformance

There are a couple of easy steps that institutions can take to conform to the model above today.

1. Add Apache/ISS/Nginx redirects to existing pages and advertise these redirects instead:  
e.g. Set up a redirect from /undergraduate/courses to /study/undergraduate/courses
2. Add a CNAME record for www. to the DNS for the domain to add support for prefix less addresses:  
e.g. CNAME www.example.ac.uk. → example.ac.uk.

Then update Apache/IIS/Nginx to add an alias for the domain:  
e.g. (Apache):

ServerName institution.ac.uk ServerAlias www.institution.ac.uk

## Linked Data?

Linked Data is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. [linkeddata.org](http://linkeddata.org)

Within the higher education sector there is a gradual movement towards [providing linked data](#) for an institution. A key part of this is understanding both what data is available, and how to present it in such a way that it's easily linked to the rest of the content on the Internet.

Linked data itself doesn't have to be housed separately from an institution's primary web presence, indeed in many cases it would be preferable to reduce duplication by providing a single, easily referenced point from which to access a resource. For this reason we've designed our recommended data model to represent a University's resources in a manner similar to linked data resources across the web.

### Unlock Data's Potential

To help unlock the potential of all your institution's data we *strongly recommend* that wherever possible you include machine readable alternatives. These should be located at the same resource name as the 'normal' resource, but suffixed with a machine readable format name. For example, /courses could (and indeed should) be expressed as [XML](#) at /courses.xml, [JSON](#) at /courses.json and [XCRI](#) at /courses.xml.