The web ecology of the online state: the case of Australia

Associate Professor Paul Henman

School of Social Science

University of Queensland QLD 4072

Australia

p.henman@uq.edu.au

+61 7 3365 7652

Dr Tim Graham
Research School of Social Science &

Research School of Computer Science

Australian National University

Acton ACT 2601

timothy.graham@anu.edu.au

Paul Henman is Associate Professor for Digital Sociology and Social Policy at the University of Queensland. His work examines the nexus between digital technology, public administration and social policy. His recent publications include *Performing the State* (Routledge 2018) and *Governing Electronically* (Palgrave 2010).

Dr Tim Graham is Postdoctoral Research Fellow at the Australian National University, Australia, with a joint appointment in sociology and computer science. Broadly, his research combines social theory and computationally-intensive approaches to analysing, understanding, and predicting social phenomena.

The web ecology of the online state: the case of Australia

Introduction

The online state consists of websites, social media accounts and apps administered by government. By state, we refer to the socio-political institution of government operating as a rational-legal form of authority via multiple agencies and actors (c.f. Scruton, 2007, pp. 662-663). The web presence of this online state comprises websites owned and managed by government agencies and institutions for innumerable purposes, which themselves are hyperlinked to other government, commercial and organisational websites. This heterogeneous network constitute a web ecology in which the online state is situated. Yet, this web ecology of the online state its shape and structure, and its relationship to the offline state is not well understood. This paper addresses this limitation by systematically examining the web ecology of the Australian online state.

Adopting a web ontology, where the presence of the state is encountered through its websites and hyperlink flows into, through and out of them, a number of key questions arise. What is the shape of the online state? What is the nature of the networks in which the online state is located? What are the most important sites in these networks? Are those networks characterised by jurisdictional characteristics, by institutional or constitutional relationships, as government networks, or by shared policy/service networks? Are these networks largely of government websites, commercial or a mix? How does the online machinery of government compare with the traditional offline machinery of government (Hogwood, 1997; Verhoest & Bouckaert, 2005)? It is these questions that this paper seeks to address.

This paper advance our understanding of the web ecology of the online state by examining the hyperlink network of government websites and their neighbours. In doing so, the paper first develops a conceptual understanding of hyperlink networks as constituting a web ecology of the online state and reviews relevant past literature. It then explains how the hyperlink network of the Australian online state was created and analysed using various digital methods and network analysis techniques. Such techniques seek to provide insight into the nature of the sub-networks (communities or modules) and how they may reflect structural aspects of the offline state. The paper concludes with a discussion of what the findings mean to the understanding of the online state in a wider web ecology.

Conceptualising the online state as a web ecology

For the purposes of this paper, the online state refers to the web presence of government institutions and agencies as evidenced by the websites owned and operated by the state. Typically each government agency will own and operate at least one website reflective of that agency and its activities and operations. There are also typically websites for whole of government or joined-up service delivery that incorporate information and/or services from multiple agencies. This include whole of government web portals, open data platforms and log-in service delivery sites. Just as a state has a portfolio of assets, including land, buildings and infrastructure, it also has a web portfolio or web infrastructure.

The online State constituted by this web infrastructure is a network. It is, in fact, a small sub-network of the world wide web, which is fundamentally constituted as a network of hyperlinks between websites (or more specifically, webpages or URLs) (Helmond, 2013, p. 3). Indeed, hyperlinks and websites (and webpages) are the fundamental building blocks of the web. In

network terminology, hyperlinks are directed edges to (webpage nodes of) a website. Indeed, websites themselves are also networks of hyperlinks between a set of webpage nodes that share a common domain name.

In building government websites, decisions are made by website developers about what other websites to link to. 'Hyperlinking is a strategic action with a hyperlink conferring information authority (source credibility) on an alternative website's informational content' (McNutt, 2010, p. 924). There are, however

> a variety of motivations for linking, including signposting valuable content, reflecting the structures and preferences of the linking organisation, and wholly functional link intended for the delivery of the site's own content rather than reflecting any kind of extra-organisational link (Nicholls, 2016, p. 161).

For example, a whole of government portal would be expected to connect to all other government agencies' websites as its purpose it to act as a knowledgeable gateway to government on the web. One could also expect webmasters creating hyperlinks to government websites in other countries (particularly for a foreign affairs website) and to other jurisdictions (namely, national, provincial or local government). Government websites will also link to non-government websites. For example, each website is likely to functionally link to their social media accounts hosted by Twitter, Facebook, YouTube, etcetera. Social service agencies, such as health and education, might also be expected to link to commercial and not-for-profit websites associated with health and education, particularly when such services are delivered by commercial or not-for-profit sectors. This network of links from government websites reflect the way government actors (as webmasters) have created an online State, a network of websites that are connected by purpose, relationship and infrastructural requirements. But this online State is only one part of the network in which the online State resides.

The online state is also the object of non-state websites linking to it. For example, mass media sites might link to Ministerial media releases hosted on Minister's government websites. Political party websites could link to their elected candidates' online state presence, typically hosted on a parliamentary website. Non-state health, welfare and education agencies could link to information about government health, welfare and education policy, benefits and services on government websites. In these cases, government websites are linked as sources of information or transaction, creating the online state as a central or nodal point in a network (c.f. Hood & Margetts, 2007).

This hyperlink network of websites interlinking with the online state can be helpfully characterised as a web ecology, a network or system of diverse website actors each in relation to other websites forming a level of inter-dependency and relationality (c.f. Henman & Graham, under review). Ontologically, this web ecology consists of three types of phenomena. First is the set of websites as entities or actors (analogous to an environmental ecology made up of plants and animals). Second is the hyperlinks as defined relationships between these actors. This is analogous to the mechanisms by which one entity can interact with another entity, for example, organs for reproduction, mouths for consumption and claws for catching prey. These two dimensions constitute the infrastructure or architecture of the web ecology. It defines what is possible. It is not dynamic. Thus, the third dimension is the dynamics of these two, the movement or web surfing by users from website to website (analogous to an entity mating with, or eating another). The first two dimensions are the space of the web ecology, the second are the flows.

When viewed in this way, a range of considerations arise about the specific nature of the web ecology. Noting that all actors are not similar, some are government (e.g. .gov), and some commercial (e.g. com). They belong to different countries (e.g. .uk, .au, .de). These can be conceived of different sub-systems in the web ecology. Each sub-system also consists of smaller sub-systems. For example, within government websites there are ones for different countries, and also within different countries there are ones for different national, provincial and local government jurisdictions. Cross-cutting sub-systems exists, such that there is likely to be a range of government, commercial, not-for-profit and international websites relating to thematic areas, such as environment, defence, health and education.

A small body of research has investigated the web ecology of the online state. One approach is to understand government in the world wide web constellation. Early work in 2005 by found that of the websites in the Australian web domain (.au), only two percent were government sites, but they accounted for 5.5 percent of all webpages (Ackland, Spink, & Bailey, 2007).

This work does not provide information about the online state within a network context. Work at the Oxford Internet Institute (Escher, Margetts, Petricek, & Cox, 2006; Petricek, Escher, Cox, & Margetts, 2006) undertook early hyperlink analysis of standalone government websites. Escher et al (2006) found that over half of inlinks to government foreign office websites (for UK, USA and Australia) were from commercial sites, while 10 to 20 percent were from government websites from the same country. In comparison, Petricek et al (2006) examined audit office websites from five countries and found only 30% of inlinks from commercial sites, and a higher proportion (around 25%) from government sites. This later study also examined outlinks from government (audit office) websites, and identified that outlinking behaviour gives preference to other government websites (about half of all outlinks) with also a high

proportion of linkages to other audit office websites, suggesting a strong linkage pattern along service domains lines.

Other research has sought to understand how the online state is structured. In short, this research seeks to find if there are patterns in the connections of government websites, and what they might reflect. Do those patterns reflect offline patterns of socio-organisational structure (such as reporting hierarchy or jurisdictional boundaries), geography, or shared policy or service interests? In a crawl of all 1077 websites of the USA Federal government, Whalen found that 'the .gov domain is structured as a hierarchical tree with centralized sites surrounded by tiers of subordinate sites' (2011, p. 11). Using linkage metrics, he also found that websites associated with one Federal agency or policy domain had strong interlinkages, but typically more linkages to other domains, showing that websites were not generally siloed within their own domain. Whalen concluded that there about a third of policy/service domains show high levels of networking with other domains (e.g. Education, Veteran Affairs, Agriculture), about a fifth were highly siloed (Justice, Defense), while about a half were in between (e.g. Commerce, Treasury, Transport). This study, however, only looked at the online state at one level of government (i.e. Federal), and does not take into account the wider network that the online state is located within.

The work of Henman et al (2014a) did this by analysing the UK's national, regional and local governments' web presence within the wider web ecology. Using a community detection method to cluster websites by 'information' flows along hyperlinks, they found that evidence of clustering of websites by policy/service domain (e.g. health), by jurisdiction (e.g. Scotland, London), by individual government agencies (e.g. Treasury) and by web functions (e.g. Facebook, Google). Notably, the most important cluster was information provision, including

data.gov.uk, Wordpress, Wikipedia, *The Guardian* and Reddit, and the second cluster was social media, thereby demonstrating that government is not dominant to the information infrastructure operating at the centre of the web ecology of the online UK state.

Some studies instead have discovered structuring of government website connections along geographical lines, particularly at local government level in Finland (Holmberg & Thelwall, 2008) and England (Nicholls, 2016). Interestingly, the latter study found no evidence of linking by service domain, perhaps suggesting that websites of local authorities have different linkage patterns than national or regional level governments. These studies also suggested linking structures along hierarchical institutional lines, where smaller parish councils are linked to their parent district councils.

Another approach is to ask how important is government in broad networks in which government would be expected to be located? Work by McNutt (2010, 2012) and Henman et al (2014b) focuses attention on the 'virtual policy networks' (McNutt 2010) of particular policy domains to identify differences in the makeup and structures within these networks. Examining five different policy domains in the USA, McNutt (2010; 2012) found that the online state had a stronger centrality or nodality (as measured by proportion of total inlinks) in the areas of Federal-provincial funding equalisation and climate change, and very low nodality in agriculture and health, which was partly explained by being a policy domain connected to international policy networks. Henman et al (2014b) analysed UK's foreign affairs, health and education policy hyperlink networks using Kleinberg's HITS algorithm and found that while key government websites had high authority in the network[1], so too did leading social media sites. On the other hand, sites with high hub scores – that is, strong outlinking behaviour – were

largely publishing or media websites (such as WordPress and the BBC), but the National Health Service was also strong in this regard in all three policy networks.

Some general observations about the online state can be discerned from this literature. Variations in their findings may be attributable to different countries and tier of government studied. Additionally, variations may also be attributable to the different methodologies used, including the way the data is generated, and what websites are included, how the network is represented, and the measures used to assess the data.[2]

Data collection: Generating the network data

The Australian state consists of a national Federal government, eight states and territories, and 546 local government jurisdictions (www.regional.gov.au/local/). The total web presence of this entire Australian online state includes multiple, perhaps hundreds, of websites in each of the Federal and State jurisdictions. Web crawling this entire web presence and their neighbouring websites would be a massive undertaking. Accordingly, in order to generate a hyperlink network of the Australian online state a set of 75 seed sites were used to start the web crawling and hyperlink generation process (see Appendix). These seed sites were identified by first selecting three of the eight states and territories (Victoria, Queensland and Tasmania) to reflect variations in population and geographical size, and four (two rural and two metropolitan) local governments in each of these three states. In each of these 16 jurisdictions we included the home portal for that government and a sample of websites for major central and service government departments covering a range of policy areas: parliament; business portal; prime minister/premier department; treasury/finance; foreign affairs; defence; health; social security; housing; education; human services; environment; law; police.[3] In cases where

these domains maintained more than one major website, we also included other key websites. For example, at the Federal level for environment we included the Department of Sustainability, Environment, Water, Population and Communities website, as well as the websites for the Australian Antarctic Division, the Great Barrier Marine Park Authority and the Murray-Darling Basin Authority. Our approach thus extends previous work that focused only on national government agency websites (e.g. Escher et al, 2006; Whalen, 2011) to also include regional and local government (e.g. Henman et al., 2014a; Holmberg & Thelwall, 2008; Nicholls, 2016).

In late 2012 hyperlink network data were collected and assembled via an iterative process that occurred in three stages. Stage one involved webcrawling each seed site using the VOSON system (http://voson.anu.edu.au/). A maximum of 1500 webpages in each seed website was crawled. This web crawl collected both outbound links (webpages that the seed site webpages link out to) and inbound links (webpages that link in to the seed sites webpages). Inbound links were collected via the Blekko API.

Stage two involved finding the outbound links for all webpages not within seed sites. Hyperlinks between these webpages were mapped during this process. To be sure, the resulting network was initially a network of individual webpages with hyperlinks as directed edges between these webpages. It contains 78,515 distinct webpages and over 1.2 million hyperlink edges.

Stage three involved a process of grouping webpages into websites. Such pagegroups were created from the webpages in the network to reflect natural online groupings constituted as websites as indicated by the same domain name. For example webpages

www.aph.gov.au/News_and_Events and www.aph.gov.au/Parliamentary_Business were grouped into the pagegroup www.aph.gov.au. This process produced a network between websites based on domain names, rather than between individual webpages. Thus, each node in the network is a group of webpages (in a single domain name). The resulting weighted, directed network contains 15,344 websites and 171,543 hyperlink edges.[4]

To be sure, while all (or most) webpages were collected for each seed site, non-seed websites comprise of only a selection of webpages that arose from the webcrawling process. Formally, this network is defined as a 1.5-degree egonet (Ackland, 2013, p. 50) representing hyperlinks between seed sites and hyperlinks between sites directly connected as neighbours to seed sites, but not the new neighbours' neighbours.

Data analysis: analysing the hyperlink networks

In order to understand the web ecology of the online state, its makeup and structure, a range of analysis techniques were employed. Part of this network is those websites that government think it is important to link to (i.e. outlinks), and part of the network are those websites that think it important to link to government (i.e. inlinks).

To understand the makeup of this network, we undertook descriptive statistics of the websites characteristics including country code and website type as indicated by generic top-level domain (e.g. .gov, .com, .org). In order to identify the important websites in the network, we first undertook an analysis of network centrality using Kleinberg's HITS algorithm (Kleinberg, 1999) which produced two inter-related scores: Authority; and Hub. In short, a high authority
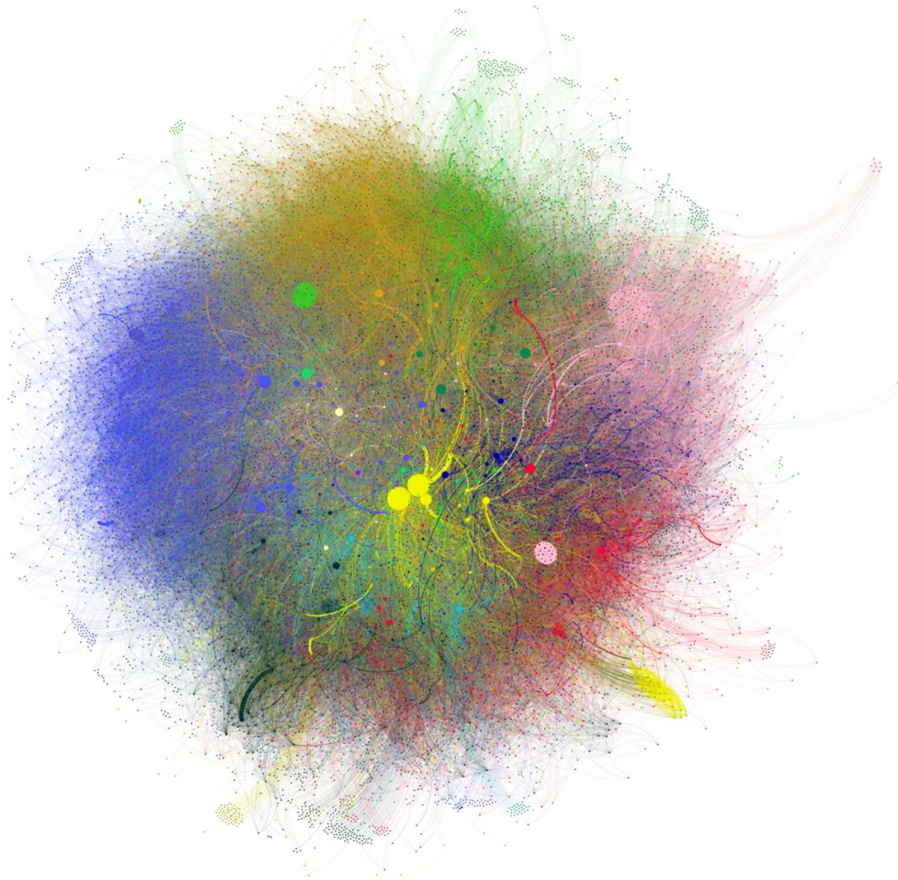
site is one in which lots of sites high hub scores tend to point, and a high hub site is one which points to a lot of high authority sites.

We also analysed the overall web ecology by identifying the natural groupings (or clusters) using the Infomap algorithm (Rosvall, Axelsson, & Bergstrom, 2009). There are other approaches to community detection in complex networks, including modularity maximization (Lancichinetti and Fortunato, 2012), Edge-Betweenness (Girvan and Newman, 2001), Fast-Greedy (Clauset et al, 2004), Multi-level (Blondel et al, 2008), Walktrap (Pons and Latapy, 2005) and BNEM (Hazef et al, 2014). Infomap was utilized not only because it is one of a few that supports directed and weighted networks, but it also scales well. Infomap identifies communities by undertaking random walks (or information flows) through the network. Communities of websites are derived by the routes through websites 'among which information flows quickly and easily' compared with the rest of the network (Rosvall and Bergstrom, 2008: 1118). This enables us to 'focus on how the structure of the extant network constrains the dynamics that can occur on that network' (Rosvall and Bergstrom, 2009: 14). We ran the algorithm 100 times, as more iterations have not been found to lead to noticeably different findings (Henman et al., 2014a).

Findings: the website ecology of the Australian online state

What is the composition of the web ecology of Australia's online state (Figure 1)? One way to answer this question is to investigate the characteristics of the various websites in the network. Of the 15,344 websites, just over half (50.7%) are commercial (.com), while government websites (.gov) make up just 14.6 percent of the network, and websites coded .org are almost a quarter (22.3%) (see Table 1). It may be surprising that a network created by crawling

government websites contains only a small proportion of government sites. However, recall that the network included websites that point to government as well. This finding is also in keeping with the composition of the entire web, which is dominated by commercial sites.

Another important consideration is the geographic composition of the network, which we measure using the country code top-level domains of the websites (Table 2). The observed frequencies of countries follows a power law distribution. Perhaps reassuringly almost two-thirds (63.8%) are Australian websites (.au). Most of the other websites in the network are 'unknown' (29.7%), which means they have no country code. This is typical of many commercial websites, such as google.com, or facebook.com, even when they may have .au versions as well. Of the remaining 6.5% of websites in the website ecology of the Australian online state, countries with major socio-cultural connections with Australia, including USA,

UK, New Zealand, Canada and many European Union countries, are predominant. Interestingly China, which is Australia's largest trading partner, is barely visible in the network (0.02%).

Table 1: Composition of the AU government network, by website type

Top Level Domain	Total network No. Sites	%	Outlinked sites No. Sites	%	Inlinking sites No. Sites	%
com	7775	50.7	2404	37.3	5919	53.2
org	3417	22.3	1549	24.1	2324	20.9
gov	2240	14.6	1802	28.0	1417	12.7
net	644	4.2	198	3.1	494	4.4
edu	428	2.8	200	3.1	352	3.2
asn	215	1.4	101	1.6	151	1.4
info	88	0.6	30	0.5	63	0.6
int	19	0.1	13	0.2	14	0.1
biz	16	0.1	2	0.0	15	0.1
nhs	5	0.03	4	0.06	1	0.01
mil	5	0.03			5	0.04
parliament	1	0.01	1	0.02		
police	1	0.01			1	0.01
other	490	3.2	135	2.1	374	3.4
Total	**15344**		**6439**		**11130**	

It makes sense to drill further into these statistics to better understand the websites to which the seed sites point and the websites that link into the seed sites. In the first case, this tells us something about what websites government regard as important to link to. Table 1 demonstrates that of the 6349 websites, the seed sites largely link to commercial (37.3%), government (28.0%) and organisational (24.1%) websites. This is suggestive of a stronger linking pattern of government websites to other government websites compared to the wider linking patterns in the full network. The countries to which the seed sites connect follows a more notable power law distribution: Australia (73.8%), unknown (21.4%).

Table 2: Composition of the AU government network, by country code

Country	Total Network No. Sites	%	Outlinked sites No. Sites	%	Inlinking sites No. Sites	%
Australia	9784	63.8	4755	73.8	6963	62.6
Unknown	4557	29.7	1376	21.4	3424	30.8
United States	182	1.2	78	1.2	130	1.2
United Kingdom	156	1.0	64	1.0	96	0.9
New Zealand	99	0.6	26	0.4	78	0.7
Germany	82	0.5	6	0.1	76	0.7
Canada	68	0.4	19	0.3	53	0.5
Italy	28	0.2	2	0.03	26	0.2
France	22	0.1	5	0.1	18	0.2
Switzerland	17	0.1	4	0.1	14	0.1
Austria	16	0.1	2	0.03	14	0.1
Spain	16	0.1	1	0.02	15	0.1
Finland	16	0.1	2	0.03	14	0.1
Sweden	16	0.1	3	0.05	13	0.1
Brazil	15	0.1	3	0.05	12	0.1
Japan	13	0.1	5	0.1	9	0.1
Denmark	12	0.1	5	0.1	7	0.1
Ireland	12	0.1	4	0.1	9	0.1
Netherlands	11	0.1	-		9	0.1
Singapore	10	0.1	1	0.02	9	0.1
Tuvalu	10	0.1	-		10	0.1
Antarctica	6	0.04	4	0.1	4	0.04
Cocos Islands	5	0.03	3	0.05	4	0.04
China	3	0.02	3	0.05	-	
Europe	9	0.1	1	0.02	9	0.1
Other	179	1.2	67	1.0	114	1.0
Total	**15344**		**6439**		**11130**	

We examine the 11,130 websites that link into our Australian government seed sites to highlight the websites that think it is important to link to the Australian government. Following a similar pattern to the overall web ecology, they are largely commercial (53.2%), organisational (20.9%) and government (12.7%) websites, and they are largely Australian (62.6%) and unknown (30.8%) country codes.

What are the most important websites in the web ecology of the Australian online state, and how important are government websites? To address this question we used Kleinberg's HITS algorithm to assess the important websites in terms of authority and hub scores, which are a value from 0.0 to 1.0, with the higher score being of higher importance. Table 3 presents the top 20 websites by authority and hub scores.[5] Authority scores indicate the value of a website's information within the network. Notably, three major social media sites – Facebook, Twitter and YouTube – have the highest authority scores. While it may seem odd that government websites are not central in this government derived network, it is worth observing that most government agencies and websites will have social media accounts and link to them from their homepages and hence social media are central for the online state. Another significant observation is that Australian federal government websites have higher authority scores than state or local government websites, even though sites from all three tiers were crawled. This demonstrates that the central authority of the offline federal government is replicated online. The major policy fields of health (#4), social welfare (#6), foreign affairs and trade (#7), taxation (#8) and environment (#9) are of high importance, and indeed health is higher than the Australian government's webportal (#5). Only two of the eight state and territory jurisdictions are listed in the top 20 sites: Victoria (#14, 20) and Northern Territory (#17). It is fascinating to observe that health Victoria is higher than the Victorian webportal, replicating the federal ranking. It is also interesting to observe the high importance given to the Northern Territory, which has the smallest population of all Australian states and Territories, especially since it was not a seed site.

Hub scores indicate the value of a website's outlinks for other websites in the network. The Australian government's webportal is the most significant hub website, which is consistent with its task to link to other parts of the online state. The webportals of the Victorian and

Queensland state governments are also in the top 20 (#10, 12). Other major informational sites rank highly, including the Australian Bureau of Statistics (#4) and the public Australian Broadcasting Commission (ABC, #7). Commercial informational sites include Wikipedia (#2) and Ask (#5) as is the blogging platform WordPress (#3). Notably nine of the top 20 hub sites are university websites, thus providing strong outward linkages to government and other authoritative websites.

Table 3: Kleinberg's top ranked websites in the AU government network

Rank	Website	Authority	Website	HUB
1	facebook.com	1.000	australia.gov.au	1.000
2	twitter.com	0.968	wikipedia.org	0.799
3	youtube.com	0.597	wordpress.com	0.650
4	health.gov.au	0.509	abs.gov.au	0.650
5	australia.gov.au	0.478	ask.com	0.646
6	centrelink.gov.au	0.474	adelaide.edu.au	0.609
7	dfat.gov.au	0.464	abc.net.au	0.595
8	ato.gov.au	0.453	csu.edu.au	0.585
9	environment.gov.au	0.423	uq.edu.au	0.580
10	abs.gov.au	0.411	vic.edu.au	0.555
11	aph.gov.au	0.405	aph.gov.au	0.553
12	abc.net.au	0.400	qld.gov.au	0.551
13	deewr.gov.au	0.393	anu.edu.au	0.544
14	health.vic.gov.au	0.363	foolkit.com.au	0.525
15	google.com	0.359	ecu.edu.au	0.521
16	immi.gov.au	0.333	newcastle.edu.au	0.491
17	nt.gov.au	0.314	swinburne.edu.au	0.491
18	business.gov.au	0.302	unimelb.edu.au	0.459
19	ag.gov.au	0.298	health.vic.gov.au	0.458
20	vic.gov.au	0.293	psnews.com.au	0.455

The above analysis provides important insights into the makeup of the web ecology of the Australian online state. This large-scale aggregate analysis does not identify the sub-systems and natural groupings present within the web ecology. Just as in an environmental ecology there are groupings around species and carbon exchange, an important consideration is whether the online groupings are characterised by jurisdictional characteristics, institutional or constitutional relationships, government networks, or shared policy/service networks. As

discussed previously, the Infomap algorithm was used to identify the community clusters of the web ecology of the Australian online state.

Table 4: Top 25 website communities based on information flow using InfoMap algorithm

Rank	Community	Primary website	Aggregated flow volume	Number of websites
1	Queensland Government	qld.gov.au	0.1138	1064
2	Twitter / Ombudsmen	twitter.com	0.0731	159
3	Victorian Government	vic.gov.au	0.0610	912
4	Health	health.gov.au	0.0523	981
5	Tasmanian Government	tas.gov.au	0.0459	587
6	Legislature	comlaw.gov.au	0.0457	398
7	Human Services 1	humanservices.gov.au	0.0447	159
8	YouTube / Google	youtube.com	0.0422	106
9	Environment	environment.gov.au	0.0310	472
10	Business / Finance	ato.gov.au	0.0289	516
11	Inter(national)	gc.ca	0.0241	567
12	Education / Employment	deewr.gov.au	0.0204	291
13	Foreign Affairs / Trade	dfat.gov.au	0.0197	292
14	Facebook	facebook.com	0.0179	8
15	Federal Government	australia.gov.au	0.0167	555
16	Health (NGOs Victoria)	health.vic.gov.au	0.0156	413
17	Media 1	doubleclick.net	0.0144	51
18	Wikipedia / Infosphere	wikipedia.org	0.0114	213
19	Education (Victoria)	education.vic.gov.au	0.0112	181
20	Media 2	news.com.au	0.0108	53
21	Legal / University	auslii.edu.au	0.0098	168
22	Defence	defence.gov.au	0.0084	145
23	Open Data	abs.gov.au	0.0083	46
24	Migration	immi.gov.au	0.0079	27
25	Human Services 2	fahcsia.gov.au	0.0076	101

Infomap identified 905 communities, ordered by total flow volume within each community. To be sure, this flow volume is not based on actual users of websites and their navigation along hyperlinks, but flow volume based on the structural characteristics of the network designated by strength and direction of hyperlinks. As with the Web, where a relatively small number of popular websites account for a majority of the world's internet traffic, so it is with the simulated flow of the web ecology of the Australian online state. The top 25 (or 2.7%) communities constitute 74.3 per cent of total flow volume, although these communities do include just over

half (54.8%) of all websites in the network. The top 5 (or 0.6%) communities constitute 35.0 percent of total flow involving about a quarter of total websites.

Table 4 lists the top 25 communities, based on flow volume within each community (loosely interpreted as importance). The table also lists the most dominant site in each community and the number of websites (or nodes) in each community. Each community has been named to reflect the makeup of that community. While communities often contain over a hundred websites it is often possible to identify a theme in each community by taking account the nature of the websites, focusing particularly on the top dozen or so that contribute the most information flow within that community. Consideration was given to the policy and service domain of the websites, the tier of government the sites are associated with, and other purposes of that site (such as social media, mass media, search engine and commercial).

Overall, we observe five different types of clustering of websites: jurisdictional; policy/service; machinery of government; functional services; and sectors. While there is often a strong theme, sometimes there are two themes as well as clustering along jurisdictional and policy/service characteristics together.

Table 5: Top 25 communities by presence of websites' government jurisdiction

	Community	Federal	ACT	NSW	NT	QLD	SA	Tas	VIC	WA
1	Queensland Government	9	0	0	0	74	0	0	3	0
2	Twitter / Ombudsmen	9	0	0	0	3	1	0	4	0
3	Victorian Government	2	0	2	0	0	0	0	157	1
4	Health	57	0	17	3	10	5	2	9	8
5	Tasmanian Government	1	0	0	0	0	0	43	1	0
6	Legislature	62	0	0	0	5	0	1	7	0
7	Human Services 1	13	0	1	0	3	0	0	3	0
8	YouTube / Google	1	0	0	1	1	0	0	1	0
9	Environment	33	0	1	0	3	0	0	2	0
10	Business / Finance	44	1	24	2	18	2	2	15	9
11	Inter(national)	0	1	0	0	1	0	0	1	0
12	Education / Employment	37	1	1	0	3	2	0	5	0
13	Foreign Affairs / Trade	21	0	0	0	0	0	0	1	1
14	Facebook	0	0	0	0	1	0	0	0	0
15	Federal Government	14	0	3	0	3	0	0	1	0
16	Health (NGOs Victoria)	2	0	1	0	0	1	1	27	0
17	Media 1	0	0	0	0	0	0	0	1	0
18	Wikipedia / Infosphere	0	0	0	0	1	1	1	2	0
19	Education (Victoria)	0	0	0	0	0	0	0	20	0
20	Media 2	0	0	0	0	2	1	0	2	0
21	Legal / University	10	1	0	0	4	1	1	9	2
22	Defence	12	0	1	1	0	0	0	2	0
23	Open Data	4	0	1	0	1	0	0	2	0
24	Migration	4	0	1	0	0	1	0	0	0
25	Human Services 2	8	2	2	0	2	0	2	1	4

Jurisdictional clustering occurs in four of the top 25 communities. Consider the first community, *Queensland Government*. The total flow volume within this community consists of 11.4% of the flow volume for the entire network. This community contains 1,064 of the 15,344 websites in the entire network. The Queensland government webportal,

www.qld.gov.au, is the highest flow volume site in this community. All but one of the first 50 websites in this community are websites of the Queensland government (including www.translink.com.au and www.queenslandrail.com.au, Queensland government's public transport sites). Queensland local council websites appear outside of the top 50. This theme is also repeated in community 3 (*Victorian Government*), and 5 (*Tasmanian Government*). Community 15 (*Federal Government*) has a similar structure, but a range of key Federal government agencies are located in other communities that appear to be themed by service domain (see below). Table 5 reinforces this jurisdictional focus by listing the number of government websites by their jurisdiction for each of the top 25 communities. For example, of government websites in the *Queensland Government* community, 74 are of the Queensland government, nine are Federal and three are Victorian, and remainder zero. Two other top 25 communities show jurisdictional clustering, but also by policy/service domain.

Clustering occurs by policy/service domain in close to half (11) of the top 25 communities. Consider *Health* (community 4). While the top five websites are health-related sites of the Australian government (i.e. Department of Health; Australian Institute of Health and Welfare, a government health information website, the health and medical research council and the private health insurance ombudsman), State government health websites are not far behind (NSW at 6; Western Australian at 9; ACT 11; NT 12; South Australia 15). Interestingly, several university websites appear in the top 30 websites in this community, higher than the key medical professional bodies (such as the Royal Australian College of GPs at 26 and the Australian Medical Association at 30). Health consumer advocacy groups, such as alzheimers.org.au and diabetesaustralia.com.au also appear in the top 50. Similar policy/service clustering including government websites across jurisdictions and non-government and commercial websites, is *Human Services 1* (7), *Environment* (9),

Education/Employment (12), *Foreign Affairs and Trade* (13), *Defence* (22), *Migration* (24) and *Human Services 2* (25). *Business/Finance* (10) has a strong showing of government agencies advancing and regulating business, despite the top showing of the Australian Taxation Office and the national Treasury as first and third websites. While these service communities have a high presence of government sites, *Health (NGOs Victoria)* at 16, consists largely of non-government health consumer and health professional organisations, as reflected in the 43% of .org websites in the community (see Table 6), alongside Victorian government health agencies. *Education (Victoria)* at 19, has a similar shape. Notably both these latter ones have a jurisdictional-geographical focus not as evident in the other policy/service communities.

The machinery of government is a third theme, evident in three communities. The *Legislature* *c*ommunity (6) includes a diverse mix of websites associated with government, law and democratic processes, rather than a specific policy/service domain. The top website, www.comlaw.gov.au, is the portal for Australian Federal government legislation, this is followed by the website for the Australian Parliament, the national Attorney-General, the national Department of Finance, and Department of Prime Minister and Cabinet. The parliamentary websites for Australian States and some other countries (e.g. UK, New Zealand, Italy) are also included, as well as the Australian Electoral Commission. Machinery of government clustering is evident in ombudsmen half of community 2 (*Twitter/Ombudsmen*), which encompasses an enormous range of ombudsmen and tribunal websites across many sectors (e.g. energy and water, overseas students, financial services, refugee and migration, social security) at federal and state levels, and ombudsmen in other countries.[6] The international machinery of government is evident in *Inter(national)* (11), which contains non-Australian government websites in a many countries (including Canada, USA, Indonesia,

Mexico and Indonesia) at national and state/provincial government, as well as international organisations (such as the European Union, World Bank and the UN).

Table 6: Makeup of network communities by country code and generic website type

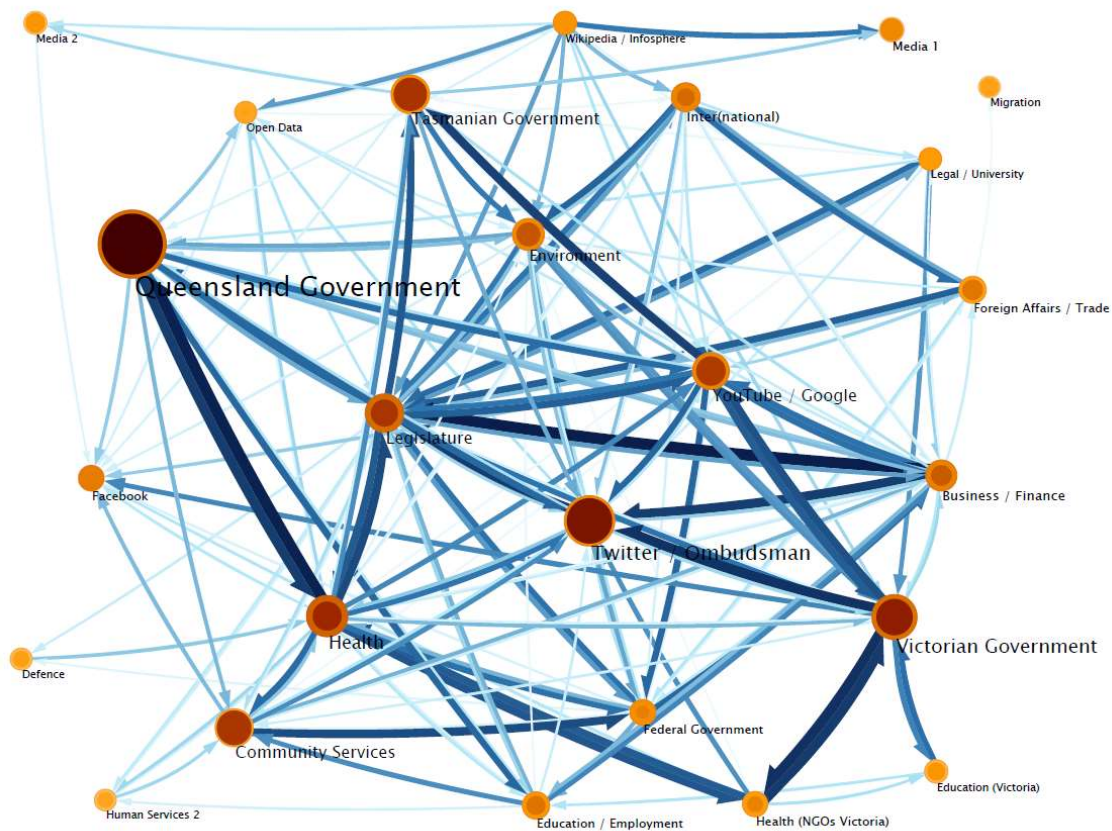|	Community	Country Code TLD (%)			Generic TLD (%)					
		AU	Other	Unknown	Gov	Org	Net	Com	Edu	Other
1	Queensland Government	78.1	2.2	19.7	20.8	25.4	3.5	45.8	1.3	3.3
2	Twitter / Ombudsmen	42.8	25.2	32.1	16.4	12.6	0.6	50.9	1.3	18.2
3	Victorian Government	81.6	0.6	17.9	28.9	18.2	4.2	45.1	0.4	3.2
4	Health	79.6	3.8	16.6	13.6	37.0	3.6	36.4	3.4	6.1
5	Tasmanian Government	77.0	1.9	21.1	26.7	16.2	2.7	50.8	0.3	3.2
6	Legislature	59.8	12.3	27.9	36.4	19.1	2.5	30.7	1.0	10.3
7	Human Services 1	79.9	1.9	18.2	13.8	13.2	5.0	63.5	3.1	1.3
8	YouTube / Google	58.5	3.8	37.7	26.4	14.2	7.5	45.3	1.9	4.7
9	Environment	64.6	4.5	30.9	21.6	27.1	4.0	40.3	1.3	5.7
10	Business / Finance	81.4	3.3	15.3	15.3	8.3	2.5	69.8	0.6	3.5
11	Inter(national)	2.8	40.7	56.4	9.9	23.6	4.9	27.7	9.0	24.9
12	Education / Employment	84.9	1.7	13.4	21.6	17.9	4.1	37.5	13.7	5.2
13	Foreign Affairs / Trade	46.2	9.9	43.8	10.3	26.4	3.8	50.3	2.4	6.8
14	Facebook	62.5	0.0	37.5	0.0	0.0	0.0	100.0	0.0	0.0
15	Federal Government	55.7	4.9	39.5	12.8	15.7	5.9	60.2	1.4	4.0
16	Health (NGOs Victoria)	74.6	2.7	22.8	8.0	43.6	2.9	37.0	2.2	6.3
17	Media 1	60.8	9.8	29.4	0.0	7.8	2.0	80.4	0.0	9.8
18	Wikipedia / Infosphere	19.3	17.4	63.4	2.3	21.1	7.5	50.7	2.3	16.0
19	Education (Victoria)	55.3	4.4	40.3	9.4	28.7	3.9	45.9	6.1	6.1
20	Media 2	83.0	1.9	15.1	1.9	5.7	0.0	90.6	0.0	1.9
21	Legal / University	82.7	1.2	16.1	17.3	19.6	4.2	47.0	7.1	4.8
22	Defence	53.8	9.7	36.6	17.9	23.4	5.5	40.0	3.4	9.7
23	Open Data	37.0	21.7	41.3	21.7	28.3	6.5	28.3	2.2	13.0
24	Migration	70.4	3.7	25.9	18.5	22.2	7.4	44.4	0.0	7.4
25	Human Services 2	85.2	1.0	13.9	21.8	30.7	3.0	38.6	0.0	5.9

Clustering occurs around functional online services in a fifth of the top 25 communities. *YouTube/Google* (community 8) has 61% of its total flow from these two web tools, but includes a scattering of other unthematized sites, including a significant showing by the Northern Territory government portal, which constitute 10% of total community flow. *FaceBook* (14) is essentially just this site. *Wikipedia/Infosphere* (18) includes Wikipedia and a raft of wiki sites (e.g. Wikimedia, Wiktionary, Wikiuniversity) as well as other major information providers, including the internet archive site (archive.org) and britannica.org. In community 2 (*Twitter/Ombudsmen*), the social media giant Twitter has an unlikely pairing with ombudsmen. The infosphere is also evident in *Open Data* (23), which includes statistics, creative commons, open source and museum sites.

Clustering also is evident around particular sectors outside of government, in three of the top 25 communities. *Media 1* (17) is dominated by websites of newspapers and related publications (principally in the Fairfax media portfolio), as well as Google's ad business, doubleclick.com, which is the community's highest flowing website. *Media 2* (20) contains the news and entertainment websites associated with Murdoch's NewsLtd corporations. Community 21 (*Legal/University*) combines two sectors. It contains many university websites as well as legal resources, including the Australasian Legal Information Institute as the top website, and some court and tribunal websites.

Figure 2 provides a visualisation of the network relationship between the top 25 communities of the web ecology of the Australian online state. It helpfully highlights the size of directional flows between communities. For example, there are large bidirectional flows between *Queensland Government* and *Health* communities, and also between *Victorian Government*

and *Health (NGOs Victoria)*, suggestive of symbiotic informational relationships. We also observe many inflows to the *Twitter/Ombudsmen* community, with minimal reciprocal outflow. *Migration* appears to have minimal amount of flow between any of the other communities, highlighting its lack of connection to other subsystems in the web ecology. In contrast, some communities, such as *Legislature*, are highly interconnected to other subsystems, reflecting the importance of this sub-system to the operation of the web ecology as a whole.

Figure 2: Network of Websites Communities of Australian Government network

Discussion and conclusion

The foregoing analysis provides a number of important insights into the nature of the online state and its location within a wider web ecology. Using only 75 Australian government seed sites (including federal, state and local government, key machinery of government sites, and major policy/service domains) resulted in a hyperlink network of over 15,000 websites and over 170,000 hyperlinks. These seed sites constitute less than half of one percent of the entire network, and link out to almost 6,500 websites with over a third being commercial, over a quarter being government, and almost three-quarters Australian. Over 11,000 websites point into these 75 seed sites, over half of which are commercial and over half are Australian based. This highlights that government's structural web relationships are strongly geographically shaped, and are dominated not by the government sector, but by the commercial sector.

This dominance of commercial sites in the web ecology of the Australian online state was highlighted when examining the most important websites in the network. The most important websites based on the information they provide to others within the network were the social media sites Facebook, Twitter and YouTube. The most important websites based on the links they provide to others within the network were the Australian Federal government's webportal, Wikipedia and WordPress. This points to the importance of commercial website platforms in providing the infrastructure for twenty-first century state (Haro-de-Rosario, Sáez-Martín, & del Carmen Caba-Pérez, 2016), compared to the classical bureaucracies and publicly owned telecommunication and transportation networks of nineteenth and twentieth century state.

We also examined the nature and structure of the web ecology's sub-systems using a community detection methodology. These online clusterings often cohered around various

characteristics, notably government jurisdictions, policy/service domains, functional online services, industry sectors, and machinery of government. Thus, online government reflects offline government in various ways, but not consistently. Our findings are different to the geographical and non-service based clustering found at local government level websites in England (Nicholls, 2016) and Finland (Holmberg & Thelwall, 2008), thereby suggesting that the online state of local government is different to that operating at national and regional levels.[7] A more direct comparison can be made with the analysis of Henman *et al* (Henman et al., 2014a) analysis of communities of the UK's government's web ecology. At a broad level, the communities in the web ecology of the Australian and British online states some similaries, including the strong presence of social media and informational websites, alongside jurisdictional and police/service clusterings.

Understanding the role, structure and function of government in the twenty-first century necessarily requires a strong consideration to government's web presence, and in particular how the online state operates within a wider web ecology. This infrastructure of the online state has not previously been examined, and it importantly highlights contemporary government's reliance on commercial website platforms for its operation and delivery of public information and services. Much more investigation is required in order to develop a picture of the contemporary state, Future research could comparatively investigate different nation states and their interconnections, the various architectural designs used by government around the world for their web portfolio, and their relative effectiveness as digital states.

Endnotes

[1] As measured by Kleinberg's algorithm, high authority are those sites to which many other sites point to, particularly those with high hub scores.

[2] Some studies only include government websites in the analysis, while others include non-government sites. Some characterise the networks as undirected, unweighted graphs (e.g. Nicholls, 2017) which represents an important loss of information compared to directed, weighted graphs. Some use network specific metrics for analysis, while other uses more crude metrics.

[3] Local authorities did not maintain separate websites for different policy or service areas, presumably because the size and scope of the authority is much smaller. Also, Australian local governments have a more limited range of functions compared to many other major developed countries.

[4] Our data generation method creates a representational bias of government seed websites compared to other websites. This is because we crawled most or all of the government seed websites, but we did not crawl the entire websites neighbouring the seeds. This bias is less significant when analysing the network at the level of websites.

[5] The initial analysis of the network produced www.myregion.gov.au as authority of 1 and www.grantslink.gov.au as hub of 1. Upon analysis it was found that these two sites had a very strong symbiotic relationship based on grants to local areas. This relationship crowded out the rest of the network and made the rest of the results unreliable. We reanalysed the network without these two websites to produce Table 3.

[6] Websites from other countries constitute a quarter of the community (see Table 6).

[7] Notably those previous studies only looked at government websites, not the wider web ecology of such sites.

References

Ackland, R., Spink, A. H., & Bailey, P. (2007). Characteristics of. au websites: An analysis of large-scale web crawl data from 2005.

Escher, T., Margetts, H., Petricek, V., & Cox, I. (2006). *Governing from the centre: comparing the nodality of digital governments*.

Haro-de-Rosario, A., Sáez-Martín, A., & del Carmen Caba-Pérez, M. (2016). Using social media to enhance citizen engagement with local government: Twitter or Facebook? *New Media & Society*, 1461444816645652.

Helmond, A. (2013). The algorithmization of the hyperlink. *Computational Culture, 3*(3).

Henman, P., Ackland, R., & Graham, T. (2014a). Networks of communities and communities of networks in online government. *Electronic Journal of e-Government, 12*(2), 117-131.

Henman, P., Ackland, R., & Graham, T. (2014b). *Social Policy on the Web: The Online Institutional Structure of Social Policy Domains in the UK.* Paper presented at the Proceedings of the 14th European Conference on e-Government (ECEG'14).

Henman, P., & Graham, T. (under review). Evaluating the Effectiveness of Government Webportals Using Online Experiments: Towards a Web Ecology Perspective. *Government Information Quarterly*.

Hogwood, B. W. (1997). The machinery of government, 1979–97. *Political studies, 45*(4), 704-715.

Holmberg, K., & Thelwall, M. (2008). Local government web sites in Finland: A geographic and webometric analysis. *Scientometrics, 79*(1), 157-169.

Hood, C. C., & Margetts, H. Z. (2007). *The tools of government in the digital age*: Palgrave Macmillan.

McNutt, K. (2010). Virtual policy networks: Where all roads lead to Rome. *Canadian Journal of Political Science/Revue canadienne de science politique, 43*(4), 915-935.

McNutt, K. (2012). From the outside in: The external face of e-government. *Journal of Information Technology & Politics, 9*(4), 319-337.

Nicholls, T. (2016). *Digital era local government in England: service reform and the Internet.* University of Oxford.

Petricek, V., Escher, T., Cox, I. J., & Margetts, H. (2006). *The web structure of e-government-developing a methodology for quantitative evaluation.* Paper presented at the Proceedings of the 15th international conference on World Wide Web.

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *The European Physical Journal-Special Topics, 178*(1), 13-23.

Scruton, R. (2007). *The Palgrave Macmillan dictionary of political thought*: Springer.

Verhoest, K., & Bouckaert, G. (2005). Machinery of government and policy capacity: The effects of specialization and coordination. *Challenges to state policy capacity: Global trends and comparative perspectives*, 92-111.

Whalen, R. (2011). The Structure of Federal eGovernment: Using hyperlinks to analyze the. gov domain.

Appendix: List of Government Websites used as Web crawl Seed Sites

AUSTRALIA - FEDERAL	
Government portal	http://www.australia.gov.au/
Parliament	http://www.aph.gov.au
Business portal	http://www.business.gov.au
Prime Minister and Cabinet	http://www.dpmc.gov.au/
Ombudsman	http://www.ombudsman.gov.au
Treasury	http://www.treasury.gov.au
Finance (Public Spending)	http://www.finance.gov.au
Data	http://www.data.gov.au
Australian Taxation Office	http://www.ato.gov.au/
Department of Foreign affairs	http://www.dfat.gov.au/
Department of Defence	http://www.defence.gov.au/
Department of Health	http://www.health.gov.au/
Australian Institute of Health and Welfare	http://www.aihw.gov.au
National eHealth Transition Authority	http://www.nehta.gov.au
Private Health Insurance Ombudsman	http://www.phio.org.au
Medicare	http://www.humanservices.gov.au/customer/dhs/medicare
HealthInsite	http://www.healthinsite.gov.au
Department of Families, Housing, Community Services and Indigenous Affairs	http://www.fahcsia.gov.au/
Centrelink	http://www.centrelink.gov.au/
Child Support Agency	http://www.csa.gov.au
Department of Human Services	http://www.humanservices.gov.au/
Department of Education, Employment and Workplace Relations	http://www.deewr.gov.au
Department of Sustainability, Environment, Water, Population and Communities	http://www.environment.gov.au/
Australian Antarctic Division	http://www.antarctica.gov.au
Great Barrier Reef Marine Park Authority	http://www.gbrmpa.gov.au

Murray-Darling Basin Authority	http://www.mdba.gov.au
Attorney-General	http://www.ag.gov.au/
Australian Federal Police	http://www.afp.gov.au
AUSTRALIA - STATE - VICTORIA	
Government portal	http://www.vic.gov.au
Parliament	http://www.parliament.vic.gov.au/
Premier and Cabinet	http://www.premier.vic.gov.au
Department of Treasury and Finance	http://www.dtf.vic.gov.au
Health	http://www.health.vic.gov.au
	http://www.betterhealth.vic.gov.au
Education	http://www.education.vic.gov.au
Environment	http://www.dse.vic.gov.au
Environmental Protection Agency	http://www.epa.vic.gov.au
Police	http://www.police.vic.gov.au
Attorney-General	http://www.justice.vic.gov.au
Community Services	http://www.dhs.vic.gov.au
Housing	http://www.housing.vic.gov.au
AUSTRALIA - STATE - QUEENSLAND	
Government portal	http://www.qld.gov.au
Parliament	http://www.parliament.qld.gov.au
Premier's Department	http://www.premiers.qld.gov.au
Treasury	http://www.treasury.qld.gov.au
Department of Health	http://www.health.qld.gov.au
Department of Education and Training	http://www.det.qld.gov.au
Education Queensland	http://www.education.qld.gov.au
Environment	http://www.derm.qld.gov.au
	http://www.climatechange.qld.gov.au
	http://www.reefwisefarming.qld.gov.au
Attorney-General	http://www.justice.qld.gov.au
Police	http://www.police.qld.gov.au
Community Services	http://www.communities.qld.gov.au
Housing	http://www.communities.qld.gov.au
AUSTRALIA - STATE - TASMANIA	
Government portal	http://www.tas.gov.au
Parliament	http://www.parliament.tas.gov.au
Premier and Cabinet	http://www.dpac.tas.gov.au
Treasury, Finance, Tax	http://www.treasury.tas.gov.au
Health	http://www.dhhs.tas.gov.au
Community Services	http://www.dhhs.tas.gov.au
Education	http://www.education.tas.gov.au
Environment	http://www.environment.tas.gov.au
Police	http://www.police.tas.gov.au

Attorney-General	http://www.justice.tas.gov.au
Housing	http://www.dhhs.tas.gov.au
AUSTRALIA - LOCAL	
Victoria – Melbourne	http://www.melbourne.tas.gov.au
Victoria – Geelong	http://www.geelongaustralia.com.au
Victoria – Mildura	http://www.mildura.vic.gov.au
Victoria – Yarra Valley	http://www.yarraranges.vic.gov.au
Queensland – Brisbane	http://www.brisbane.qld.gov.au
Queensland – Townsville	http://www.townsville.qld.gov.au
Queensland – Mount Isa	http://www.mountisa.qld.gov.au
Queensland – Roma	http://www.maranoa.qld.gov.au
Tasmania – Hobart	http://www.hobartcity.com.au
Tasmania – Launceston	http://www.launceston.tas.gov.au
Tasmania – Burnie	http://www.burnie.net
Tasmania – Queenstown	http://www.westcoast.tas.gov.au