Assessing Accessibility

Usability & Infrastructure of Open Data Portals

Umberto Boschi, Tommaso Elli, Phillip Morris, Oriane Piquer-Louis, Guillén Torres, Jeroen de Vos



Intro

The Open Data you need lives HERE



q			abert@s PORTAL OPEN	
Q		and prove		
	Search Open Baltimore		que é? o ca	tálogo condicións de uso id
	â			
	ovemment	City Services	Organiza o calendario do	
			próximo curso	
	*	< 🧱	Descarga o calendario escolar 2017-2018, cos periodos vacacionais e oddas as commemoracións e aplícao no teu ordenador ou dispositivo móbili	2200
Housing		Crime	conmemoracións e aplicao no teu ordenador ou dispositivo móbil!	
li				
	/	•		J 12345 FE
97 Gun		Financial		67890
		and the second		
	9	🐶 🛃	Catálogo de datos	Últimas aplicacións
60	graphic	Health		MeteoRoute
1			Os māis vistos Os últimos	Información meteoroló. oceanográfica
	<u>e</u>	🔺 🦾	Calendario escolar 2017-2018 Calendario CSV ODS XLSX	personalizada para rut traxectorias, tanto por t
Neg	ibarhoods	Public Safety	Ensino e formación	como por mar.
				Servizos no mapa
	عر	🙂 💽	Talleres artesáns	Localización xeográfic equipamentos e serviz públicos de Galicia:
Pub	lo Works	Transportation	Cultura e ocio	administrativos, cultura educativos, sanitarios.
		1112	CO emprego	0
- Constant	Publishing Departme		Calendario laboral 2017	
		nis	Calendario CSV ODS XLSX	Formatos máis comúns
	Aging and Retirement Education Baltimore City Department of Transportation		empresa e emprego	RDF CSV RSS
	Battimore City Health Department Battimore Housing		Relación de Postos de Traballo (RPT) do ámbito	Semantic Web Comma- Really Si Standards Separated Syndicati Values
	Baltimore Office of Promotion and the Arts Board of Elections		de Función Pública	WMS XML KML
		71	Administración pública	Web Map Extensible Keyhole Service Markup Markup
	Enoch Pratt Free Library Enterprise Geographic Information Services	,		Language Languag
			O Máis	0
			XUNTA @ Xunta de Galcia. Información mantida DE GALICIA Oficina de rexistro e información Suxesti	e publicada na internet pola Xunta de Galicia ons e queixas Aviso legal Atendémolo <u>RSS</u>
	Minority and Women-Owned Business Developmen Non-Governmental	4.1		
			政府資料開放平臺 全部資料集	互動專區▼ 最新消息 諮詢
	Public Works Recreation and Parks		DATA.GOV.TW	



Central de Dao

Este site ainda está em desenvolvimento; os recursos aquí contidos poder ao ralterados ou corrigidos a qualquer momento.

A Central de Dados é um

publicados são provenientes de diversas fontes e estão sob a

curadoria do

da Open Knowledge

objetivo é catalogar, afinar, complementar e publicar conjuntos de

informação pública, para que possam ser

Portugal. O nosso

reutilizados em visualizações, webapps, consultas e outros fins.

É possível seguir as novidades no Twitter!

A Central de Dados é uma ferramenta aberta e pública, e todo o código

para ela criado está disponível segundo uma

關於平臺▼

licenca de software livre.

登入/註冊

へ 資料集 ▼

Datas de eleições e referendos

Datas de atos eleitorais e referendos em Portugal desde 1975 Fontes: CNE MAI

repositório aberto de datasets Open Data em Portugal. Os dados aqui

Resultados das eleições legislativas desde 1976 Fontes: Tretas.org CNE MAI

Transparência Hackday e Eleições Europeias

Resultados das eleições para o Parlamento Europeu Fontes: CNE

Governos e Legislaturas

Governos, cargos e legislaturas Fontes: Parlamento.pt

Datas das sessões plenárias do Parlame

Lista das datas de todas as sessões plenárias no Parlamento des Fontes: Parlamento

Óbitos em Portugal

Vigilância de Mortalidade Fontes: Ministério da Saúde

Violência doméstica

Número de ocorrências de violência doméstica entre 2008 e 20 Fontes: DGAI Direção Geral de Administração Interna

Nomes próprios em Portugal

Listas de ocorrências de nomes próprios, e de nomes próprios a rejeitados pelo IRN Engres: IRN Público et (2014) Público et (2015) Diário de Noticias (2016

Códigos postais

Códigos postais e as áreas que lhes correspondem Fontes: CTT

SPARÊNCIA HACKDAY PORTUGAL NOS DATE WITH DATA

Portugal 📲 Journalism++ Agradecimentos 💮 Open Knowledge Internatio

Portals mediate access to data

They have their own particular discourse

Each portal makes a 'promise' to the user







Follow

It seems the web community has decided that developer experience dwarfs user experience, and I'm not sure what to do about that ...

9:32 AM - 30 Dec 2016

Research Question

How can we measure the accessibility of OpenData Portals?

(with digital methods)

- "The frontend"
 - Semantic analysis: what are portals focused on?
 - Spatial analysis: relative relevance of the different sections in the portals
 - Code intricateness: is it correlated with visual intricateness?
 - Usability.gov-style analysis.
- Online talk?
 - Is the Open Data community voicing satisfaction/dissatisfaction online?
- Infrastructure
 - Programming Issues: how messy is the Portal's infrastructure?

Seven portals analyzed (frontend and online talk)

- 1. Data.gov.uk (United Kingdom)
- 2. Data.gouv.fr (France)
- 3. Data.gov.au (Australia)
- 4. Dati.gov.it (Italy)
- 5. Datos.gob.mx (Mexico)
- 6. Data.gov (USA)
- 7. Opencanada.org (Canada)

Plus one case study about infrastructure issues → gov.uk (best rated according to openberometer.org + full source code on Github)

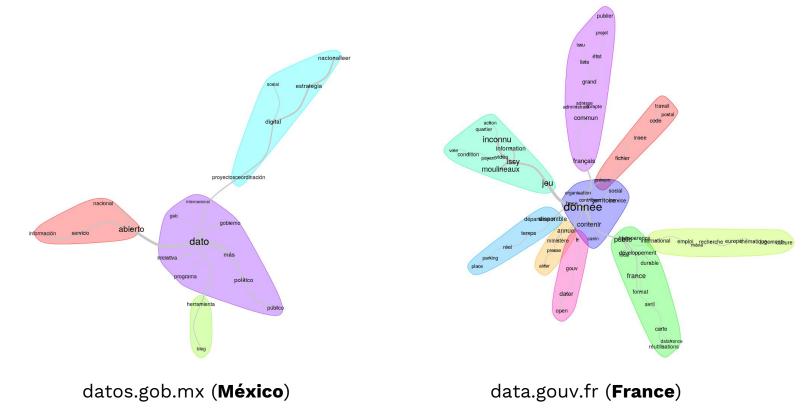
Semantic analysis

Research question : what do the portals *promise* to their readers ?

- 1. Quick semiotic analysis of the frontpage to see what is the first impression on the reader = what the website actually *means* ?
- 2. Scrape all words on the front page
- 3. Co-occurrence analysis (with community detection) with Iramuteq

Result: discourse/promise of the portal

Semantic analysis



Methods > "the frontend" (2/4) Spatial analysis

data.gouv.fr		Plateforme ouverte des données publiques françaises
Découvrez POpenDuta Données Tableau de bord		Commission / Inscription
C. Recentro Aquada et alimitatio Caha Rocate at Retroch Academia function International Engine Legranice, Recentor Martia at Respa Karel at Rout Banda Bantana, Statusch, Studien	Partagez, améliore; données publiques	z et réutilisez les
		DERNIÈRES RÉUTULISATIONS
Image: Distribution	L'annualité en derondes LES DONNÉES DES ÉLECTIONS Tous les deutais électoreux disponitées sur deu gour é	
JEUX DE DONNÉES À LA UNE DERVERS JEUX DE Mar Strong des principaises et de lours databasements (SEEX, BRET) D'SCE 0 24 = 30 Chocopage descrivéres formersent Recent des principaises d'Annonement Chocopage des addressivers formersent Recent des principaises d'Annonement Chocopage des addressivers formersent Recent des principaises d'Annonement Chocopage des addressivers formersent Recent des principaises d'Annonement Chocopage des p	Fichier des poleons - Edition 2016	CL + 4 CL + 4
tacqual like d'Operditivethip 2 3 + 20	BAN 0 22 + 21	Developpement (M43) 2915 0 1 8-1
Addes & la presea : classantant des titres preses addo 0.4 ± 4	Penarena des grands projets El de l'État (1 + 1 = 2 BYDR FLUS	Ensis officialis des codes protax 0 31 + 80
LE BLOG	7 Juin 2017 Transparence de la commande pu l'engagement de la France, du loc	al à l'international
LOPEN DATA THÉRATOLUS D'ut i gu chipe Apricion e d'arcette Brut qu chipe Apricolos e d'arcette Brut qu chipene D'ut qu chipene	NESEAU CONTACT Casawamat 2 C C C C C Anada C C C C C C C C C Anada C C C C C C C C C C C C C C C Anada C C C C C C C C C C C C C C C C C C	etalab

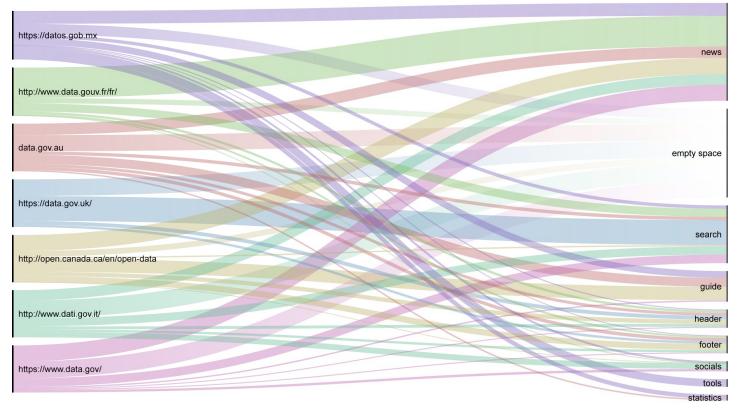
 \rightarrow

Result: How do institutions organise content across their portal home pages?

data.gouv.fr		
Diseases ToperData Dannies Terres in tert		II - O Serverse / Insortion
Autor Proceedings Application of advancement Classic Research of Bonds Advanced in Research Advanced in Advanced Depth of Design Advanced Technique Advanced Technique Research Research Advanced Technique		
MEILLEURES RÉUTILISATIONS		DERNIERES RÉUTILISATIONS
JEUX DE DONNÉES À LA UNE DETRIGENT JEUX DE		
Time form the articpline of the lower Index (1993), 1992) - N + 20	Patter des princes - Edites 2018	Canal Publication Constructions do Fermination (LEDI-7-1 da Costo da Torval) al 1 - 2 - 2
Atompanga attrictated? command Anopan data ? Operationality - 10 + 10	BAN DAY- Same Advesse Nationals	Cheffree de l'Aute Publicer au Destrongenerant (MPO) 2013
Contraction and the brackfultures including on the plane for 3 works = 1.2 × 10	Las francis publiques locates 2015	Anthographerum anterpreteine de la Cong des comptes (2004-2004 et 2014-2015)
Alders & its pressure - classesteared data Theory of pressure addition	Parameter das grande projets 11 de ribut	$\label{eq:alpha} \underbrace{ \begin{array}{c} \label{eq:alpha} \\ \begin{tabular}{lllllllllllllllllllllllllllllllllll$
	IN YOM PLUS	
LE BLOG		
Contrats commande	7 An 807 Transparence de la commande pul l'engagement de la France, du loci 1. La senerence de a comarce publice, un resultere	a a l'international
A COPER DATA THEMATIQUES	PETRAL CONTACT	atala
	Construmental Promoti Langenomenantal Construmenta Construmenta Data guardi	



Spatial analysis

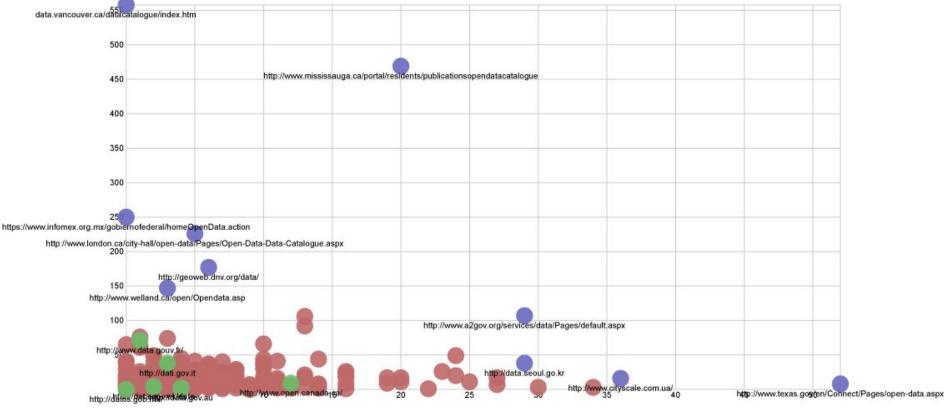


Recognised 8 different "types" of section: header, search, guide, tools, news, socials, statistics, footer.

Code analysis

- 1. Search for specific keywords on the code
- 2. Plot screen-captures according to code intricateness.

Result: Does the structure of the code relates to visual intricateness?



Information Architecture analysis

Adapted Usability.gov analysis based on Organization Structure and Organizational Scheme:

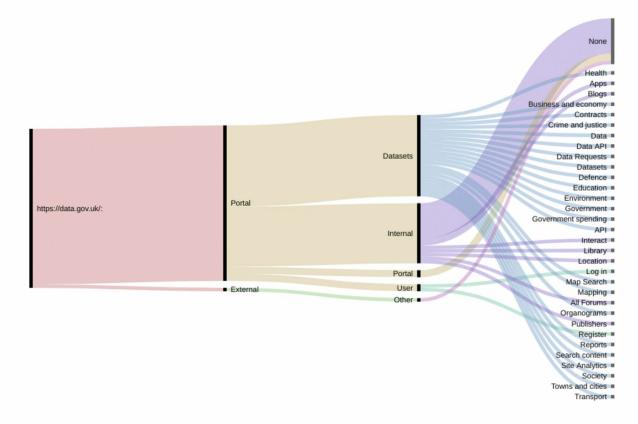
- 1. Map the structure of links present on the front page
- 2. Analyze the link arrangement on the front page

Result: Does the amount, structure and presentation of links affect data portal usability?



Information Architecture - Organization Structure analysis

Result: UK https://data.gov.uk/ 45 total links 44 internal links, 1 external links



Information Architecture - Organizational Scheme analysis

Result: Navigating by the categories given on the front page will not include most datasets

Nation	Open Data Portal	Total Links	Total Datasets	# co2 Datasets by search	Total Categories	Icons	Total Datasets in Categories
Australia	http://data.gov.au/	48	30,135	25,535	(32)	No	480
Canada	http://open.canada.ca/en/open-data	119	118,360	57,943	(19)	No	320,459
France	http://www.data.gouv.fr/	269	27,437	21	9	No	-
Italy	http://www.dati.gov.it/	127	18,222	59	13	Yes	6,675
Mexico	https://datos.gob.mx	N/A	25,804	626	11	Yes	59 (2,280)
U.S.A.	https://www.data.gov/	77	195,790	169,747	14 (20)	Yes	20,236
United Kingdom	https://data.gov.uk/	45	41,346	28,421	12(10)	No	39,136

Information Architecture - Organizational Scheme analysis

Result: The UK deemphasizes the categories of Crime & Justice and Defense

DATA.GOV.UK	Home Da	Register Log in ta Apps Interact <u>Search for data</u>	THEME	Historic Flood Map			
			Environment (12982)	Environment Agency Historic Flood Map is a GIS layer showing			
Search for data			Towns & Cities (7196)	Recorded Flood Outlines from river, the se areas of land that have			
			Mapping (5473)	areas or faild that have			
Find data published by government dep	0 11		Government (3135)				
and local authorities. You can use this da government works, carry out research of			Society (3009)	HM Land Registry Price Pai			
, , , , ,			Health (2066)	HM Land Registry Price Paid Data tracks the residential prop			
Browse data by theme			Government Spending (1852)	lodged with HM Land Registry for registra			
Business and economy	Environment	Mapping	Education (1338)	residential property			
Small businesses, industry, imports, exports and trade	Weather, flooding, rivers, air quality, geology and agriculture	Addresses, boundaries, land ownership, aerial photographs, seabed and land terrain	Business & Economy (1124)				
Crime and justice	Government	Society	Transport (1120)	GB Road Traffic Counts			
Courts, police, prisons, offenders, borders and immigration	Staff numbers and pay, local councillors and department business plans	Employment, benefits, household finances, poverty and population	★ less ★	Department for Transport The Department for Transport collects tra			
Defence Armed forces, health and safety, search and	Government spending Includes all payments by government	Towns and cities Includes housing, urban planning, leisure, waste and energy consumption	RESOURCE FORMAT	level of traffic on roads in Great Britain. Th street-level data for			
rescue	departments over £25,000		HTML (12372)				
Education Students, training, qualifications and the	Health Includes smoking, drugs, alcohol, medicine performance and hospitals	Transport Airports, roads, freight, electric vehicles, parking, buses and footpaths	CSV (7067)	Historical monthly data for			
National Curriculum			WMS (4919)	Historical monthly data for			
			WCS (3961)	Monthly Historical information for 37 UK I			
FAQ Code of conduct Accessibility Cookies po	licy Privacy Moderation policy Contact Terms and co	onditions API About	XLS (2270)	the early 1900s, but some go back as far a temperature			

Methods > "Online talk"

Online Talk and Network Analysis

How is the usage of the portals reflected on social networks (Twitter)?

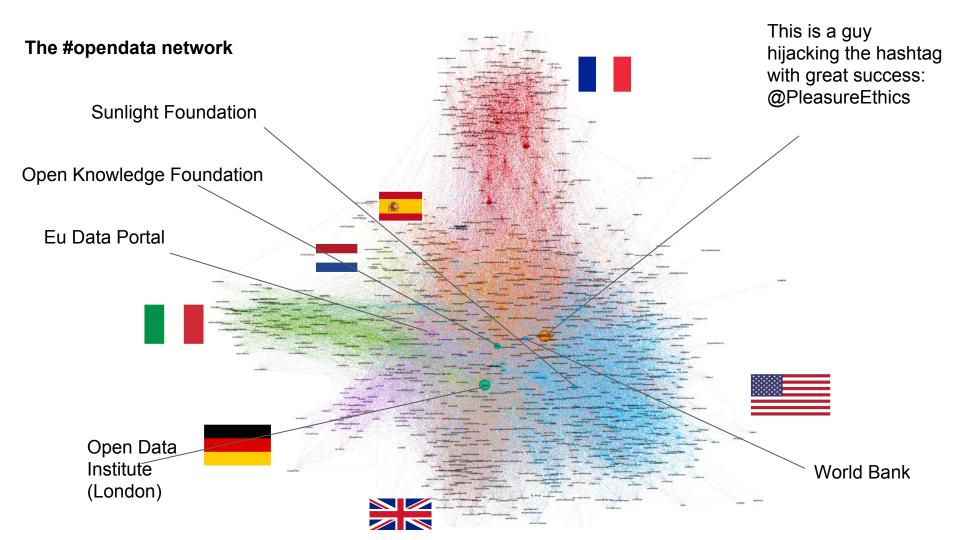
Which kind of actors are discussing open data on social networks?

Can we find evidence of data portals usage by citizens? Are there any negative experiences? (spoiler: we couldn't find any)

Online Talk and Network Analysis

Process:

- Collecting the data.
- Gephi: identifying clusters.
- Magnifying and dissecting the clusters: describing the nodes, using a sample cluster.
- Close reading: Isolating "negative" tweets and identifying dissenting voices.



shom fr pole emploi babgi toulousearchive acteurspublics ignfrance sncf infopresse assembleenat chil ecologienergie luchelot amel actu groupetaposte courdecassation dila weet lusinedigitale ti france laurentrightametrogre christianbuchel mandonthierry ercy inseeffalimag 15marches ministerecc courdescomptes ratzillas fcharles trojette ivplace conum regionepablide gfil sberne openfaw fr latribune gouvernementfr yvestyrode sabineblanc nextinpact osm fr lesechos enedis place beauvagap digital douane france metropolerennes iauidf numaparis scoppeniche mairiedigne opengovtr ciscofrance ca 94 mellille hatyp infogreffe grtgaz journaldunet france sg map openfoodfactsfr interconnectoenpaca armellegilligedardscitogeneiverdier sncfopendata lemondefr datebzh frenchweb sandrinemathon mentpelliertnavellelemaire latelier sciencespo tisemetropole opendata fr romaintalanne educationfrance datagouvfr laurelucchesi open de la company lapaillasse INCT groupekeois nedefrance archimagredad opendatasoft fr remaintales opendatagage okinfr epidemium cc sncf digital sup recherche kisiodiatioud ita la navitia paris afd france mathieucaps groupebpce opendata forum paulaforteza libertic grouperatp cfimedateham

data issy

Actors:

- Institutions
- Non-Profits
- Media
- Banks

rofeldo

- Tech companies -
- Open data municipality projects

What about the negative experiences?

- Selected from the corpus the tweets containing **sad/angry emojis**.
- Got 320 tweets against a total of 1.600.000
- A significant percentage of those are complaints about the (temporary) removal of the US data after the election of Trump.
- Most of them are people expressing **frustration** about working with data or **attending open data events**.

2015-08-20 christine_peets Drowning in a sea of #opendata (2) #dissertationfun #twomoreweeks

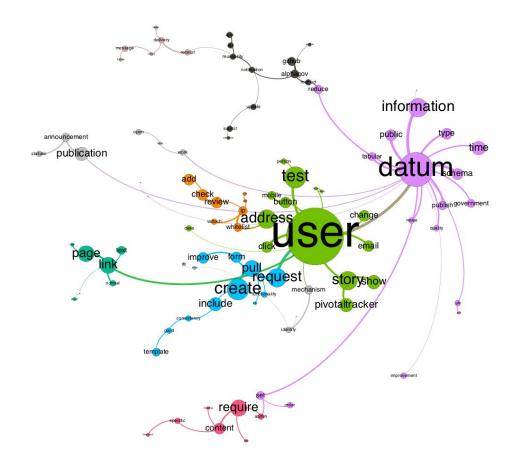
Methods > "the backend"

Issue analysis

- 1. Scraping the open/closed issues of government projects on Github related to Open Data (14 international projects, the whole gov.uk website)
- 2. Co-occurrence analysis (with community detection) with Iramuteq
- 3. Issue-Label analysis (frequency

Result: A sneak peek of what programmers discuss and consider relevant

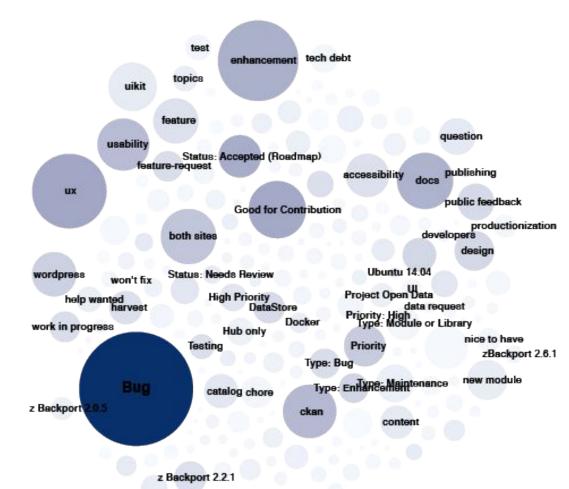
Methods > "the backend"



Issue analysis: gov.uk

Subset of issues related to **accessibility**

Methods > "the backend"



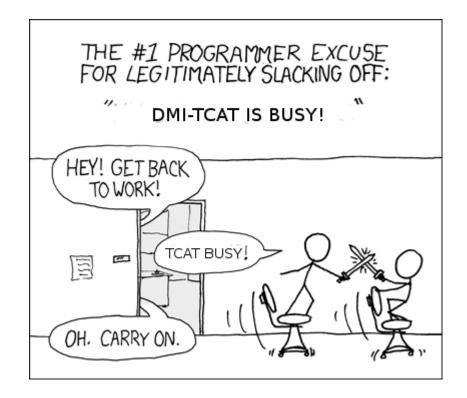
Issue analysis: gov.uk

Issue labels: frequency (size) and number of comments (color)

Conclusion

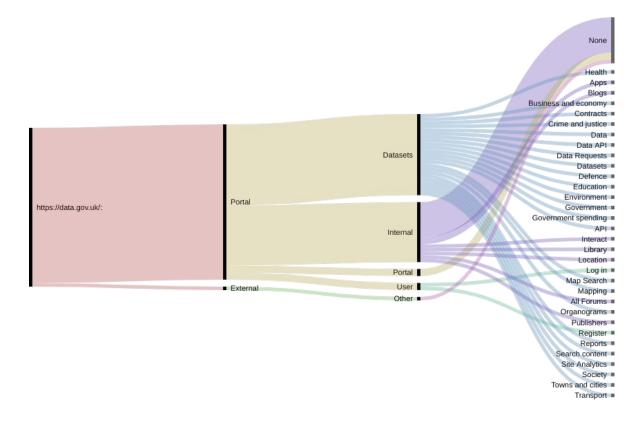
- To get a picture of the accessibility of a portal, it is necessary to combine different types of methodologies (only one doesn't make it)
- Although most portals look alike (international standards), their discourse and the type of relationship they establish with the user are **very** different.
- There seems to be a relation between how complicated the infrastructure of a portal is and its accessibility.

So long and thanks for all the fish!



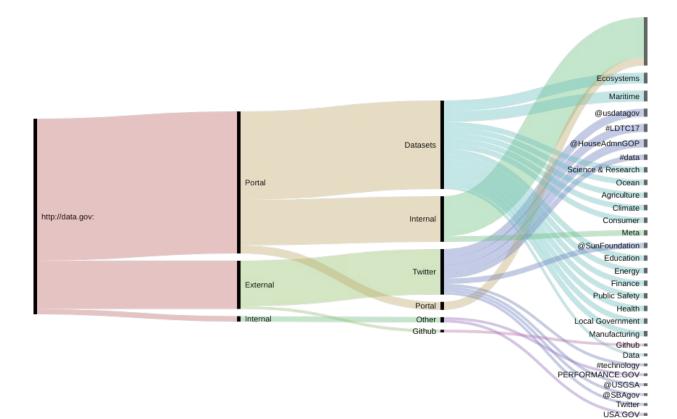
Information Architecture - Organization Structure analysis

Result: UK https://data.gov.uk/ 45 total links 44 internal links, 1 external links



Information Architecture - Organization Structure analysis

Result: U.S.A. https://www.data.gov/ 77 links 56 internal links, 21 external links



Information Architecture - Organization Structure analysis

Result: France http://www.data.gouv.fr/ 269 total links 249 internal links, 20 external links

