



مباحث فنی شفافیت

دوره آموزشی شفافیت



|  |  |
| --- | --- |
| شناسه سند | |
| موضوع | آموزشی |
| طبقه‌بندی | دوره آموزشی |
| وضعیت انتشار | منتشر شده |
| نسخه | 1 |
| تاریخ تصویب | ‏20 / 04 / 1396 |
| نویسنده / مشارکت کننده | SHAHID |

# فهرست مطالب

* وب معنایی (وب 0/3)
* رابط برنامه‌نویسی (API)
* قالب‌های رایج
* کاربردهای XML
* کیفیت داده
* داده متصل (Linked Data)

# مقدمه

همانطور که در جلسات قبل توضیح داده شد شفافیت یا ارائه اطلاعات سبب ایجاد تحلیل‌هایی می‌شود که این تحلیل‌ها، بخاطر پیگیری یا مکانیزم خودکنترلی که در شفافیت وجود دارد باعث مطالبه اصلاح رفتار شده و درنهایت این مطالبه، موجب اصلاح رفتار می‌شود.

بین ارائه اطلاعات و تحلیل اطلاعات بحث بسیار مهمی وجود دارد. فارغ از این که چه اطلاعاتی و آن اطلاعات چگونه ارائه شود، بر این‌که آیا اطلاعات قابل تحلیل هست و چقدر و چگونه می‌توان آن را تحلیل کرد، تأثیر دارد.

در تمام دوره مباحث شفافیت، ابعاد حقوقی شفافیت مورد بحث قرار می‌گیرد که چه اطلاعاتی داده و چه اطلاعاتی داده نشود.

این جلسه به ابعاد فنی بحث شفافیت می‌پردازد که اطلاعات چگونه باید داده شود؟ به‌طور کلی اگر همه محورهای اطلاعات داده شود اما اطلاعات درست نباشد، شفافیت رخ نداده است.

# API چیست؟

می‌توان گفت که API اولین و مهم‌ترین بحث از مباحث فنی می‌باشد. در وب، انتقال داده به یک سرور بر عهده API است.

شرکت MuleSoft یک شرکت بزرگ است که سرویس‌های API ارائه می‌دهد.

APIسرویسی ارائه می‌دهد که از طریق آن، هر کسی می‌تواند یکسری اطلاعات مشخص را از یک سرور خواسته و سرور آن اطلاعات را به او بدهد.

به عنوان مثال فردی می‌خواهد قیمت پروازهای مختلف را برای سایت خود به دست آورد. چون دسترسی به تمام سرورهای شرکت‌های هواپیمایی را ندارد درنتیجه شرکت‌های هواپیمایی سرویس‌های API ارائه می‌کنند که اگر کسی بخواهد از اطلاعات آن‌ها استفاده کند از آن API و قالب و شرایط استفاده کند. بنابراین از طریق فراخوانی API اطلاعات را به دست آورده و در سایت خود نمایش می‌دهد.

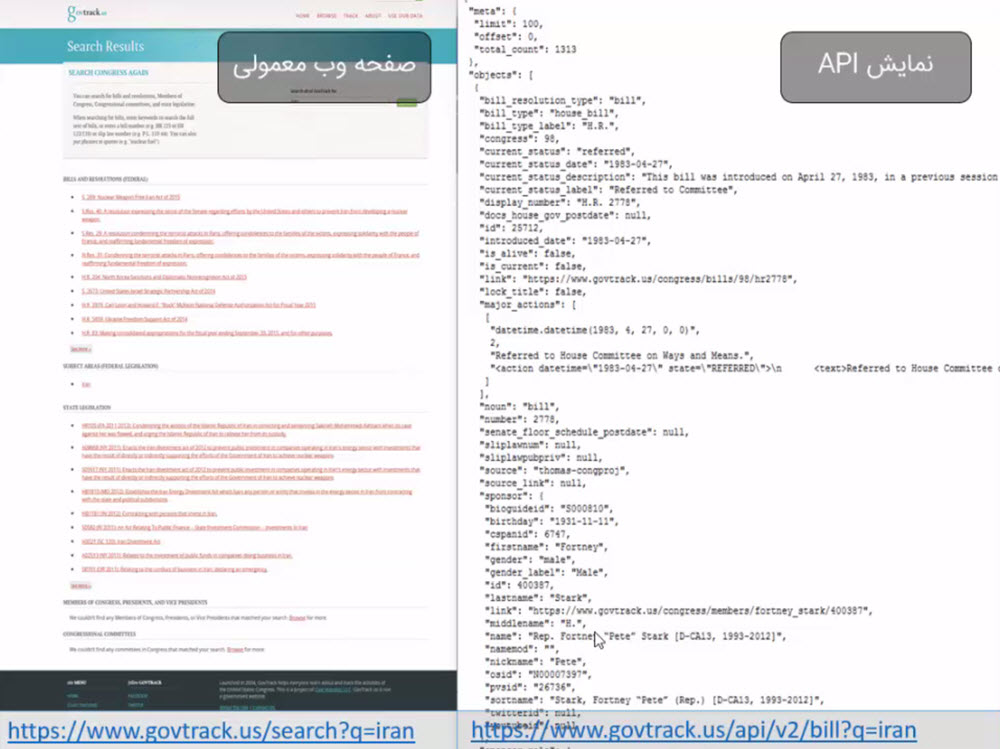
## مزایای API

قبل از API، اطلاعات داده در وب به صورت فایل معمولی بود. در اینجا ارائه اطلاعات به صورت فایل با ارائه اطلاعات از طریق API مورد مقایسه قرار می‌گیرد:

* هزینه اطلاعاتی که روی وب قرار می‌گیرد برای منشرکننده از طریق API بیشتر است اما به صورت فایل معمولی کمتر است.
* API نیاز به پشتیبانی بیشتری نسبت به فایل معمولی دارد.
* ترکیب ماشینی اطلاعات با دیگر داده‌ها از طریق API ساده است اما در فایل معمولی دشوار است.
* استفاده از API در پلتفرم‌های مختلف ساده اما در فایل معمولی دشوار است.
* عمر داده در API در لحظه (زنده) است اما در فایل معمولی دارای تأخیر (مرده) می‌باشد.

## مثال تفاوت اطلاعات به صورت وب معمولی با API

تصویر سمت چپ یک صفحه وب معمولی و سمت راست API همان صفحه را نشان می‌دهد که به صورت داده ارائه شده است.



## سایت Programmableweb

هر کسی که API را ارائه می‌کند آن را در این سایت معرفی کرده تا دیگران بدانند چنین دیتایی وجود دارد و از آن استفاده کنند. درواقع یک سایت مرجع می‌باشد. مثل همه شرکت‌های مطرح، دانشگاه‌ها، کتابخانه‌ها و ...

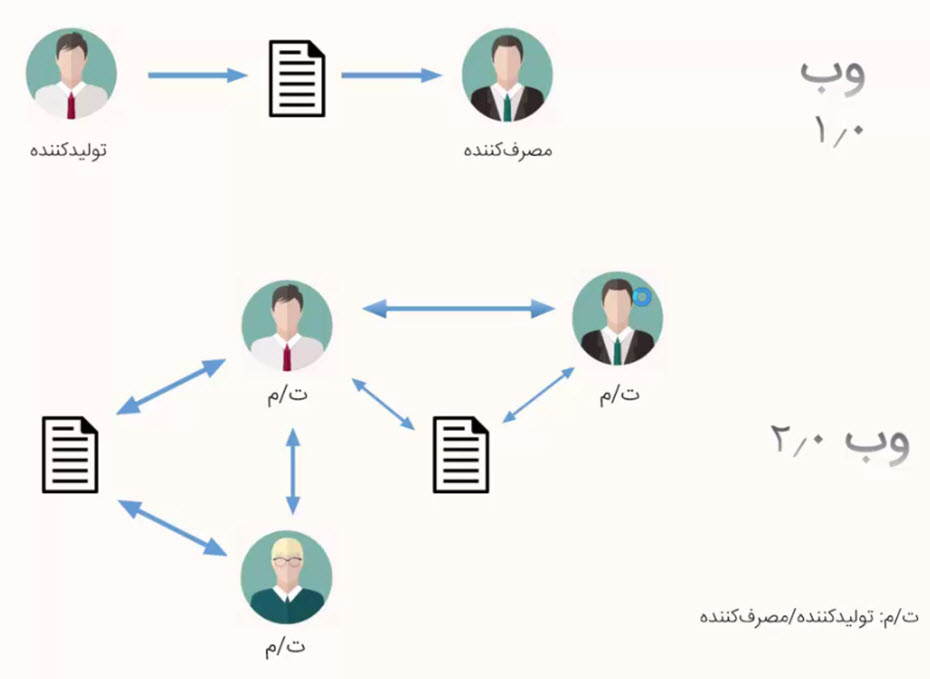
mash ups سرویس‌هایی هستند که دو یا چند API با هم ترکیب شده و API جدیدی ایجاد می‌شود.

تنها سایت ایرانی که اطلاعات را به صورت API ارائه می‌دهد سایت khamenei.ir است که می‌توان اطلاعات مربوط به آن را به صورت API دریافت کرد. در پایین صفحات این سایت، بخشی به نام API وجود دارد که در آن نوع خروجی، نوع پارامتر، نحوه دادن اطلاعات و ... توضیح داده شده است.

# وب معنایی (وب 0/3)

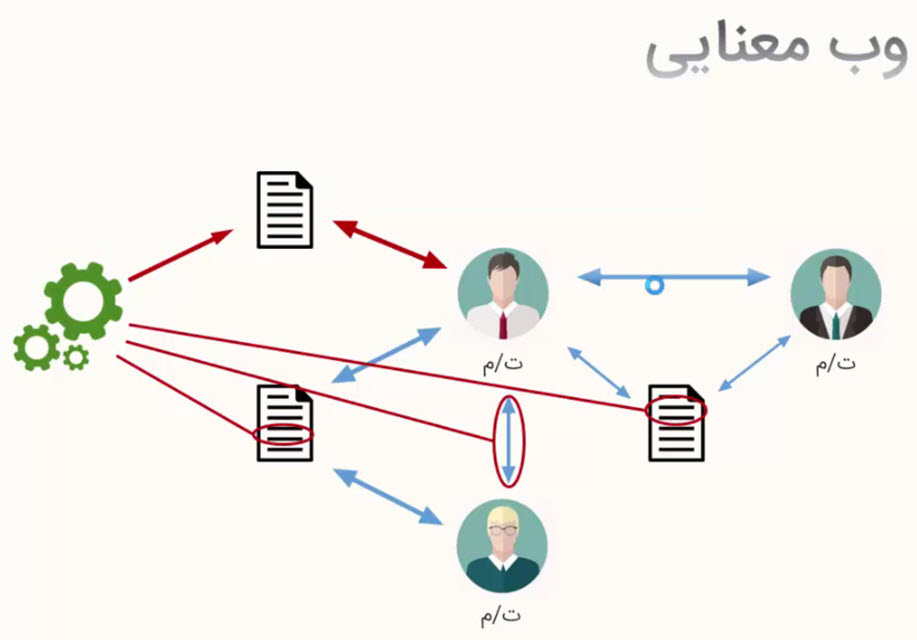
در وب 0/1 یک نفر تولیدکننده محتوا و یک نفر مصرف‌کننده است.

در وب 0/2 تولیدکننده و مصرف‌کننده با هم یکی می‌شوند.



در وب 0/3 یا وب معنایی مشارکت ماشین در تولید محتوا بسیار بالا رفته و می‌تواند هم محتوا را بخواند و هم ارتباطات را کشف و مبتنی بر آن کشف و ارتباطات، اطلاعاتی را ارائه دهد.

نکته مهم در وب 0/3 این است که موتور جستجو یا ماشین چطور باید بفهمد که متنی که در روبروی او قرار دارد چه نوع متنی است.



# قالب‌های رایج مناسب انتقال داده

بحث اصلی درباره قالبی است که اطلاعات از طریق آن‌ها منتقل می‌شود. قالب‌های جدید بخصوص در XML اطلاعات را از نظر معنا overload می‌کنند یعنی معانی را به اطلاعات می‌دهند.

## XML

زبان نشانه‌گذاری شده و توسعه‌پذیر که برای ارائه اطلاعات درباره‌ی یک سند، از برچسب‌ها (تگ) استفاده می‌کند.

ارائه اطلاعات از داده به خواننده و توصیف اسناد و داده‌ها توسط برچسب‌های XML صورت می‌گیرد. در صورتی که برای نمایش اطلاعات در مرورگر از برچسب‌های HTML استفاده می‌شود.

استفاده از XML برای خوانش اطلاعات بسیار راحت، غیرمبهم، توسعه‌پذیر، مستقل از پلتفرم و بسیار مفید برای انتقال داده از یک مکان به مکان دیگر در سرتاسر اینترنت است.

### مزایای XML

* داده را قابل حمل می‌کند
* قابل خوانش توسط انسان نیز می‌باشد
* بسیار رسا است چون معانی در کنار داده‌ها قرار دارد
* انعطاف‌پذیر و قابل سفارشی‌سازی (برچسب‌ها قابل توسعه)
* قابل استفاده برای برنامه‌ها
* تبدیل ساده با سایر قالب‌ها
* استانداردها و ابزارهای افزوده‌ی متعدد
* استفاده و پشتیبانی گسترده

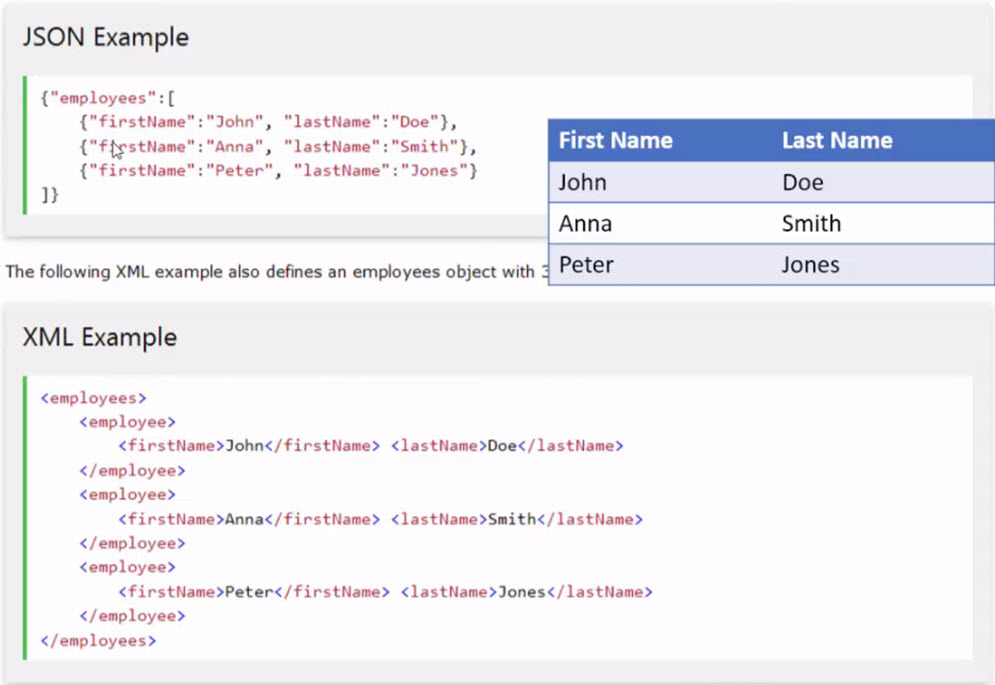
## مقایسه JSON با CSV

قالب JSON در API که فراخوانی می‌شود متداول است و خروجی از جنس JSON می‌دهد. اما در قالب CSV داده‌ها با کاما از هم جدا و داخل یک ستون نمایش داده می‌شود.



## مقایسه JSON با XML

نوع قالب در هر دو با هم تفاوت دارد.



## نمونه‌ای از کاربردهای XML

### آکوما انتوسو (Akoma Ntoso)

متن زیر بخشی از یک قانون، ماده و تبصره را نشان می‌دهد. افراد با دیدن آن متوجه موضوع می‌شوند اما ماشین از آن فقط رشته ای از کاراکترها را می‌فهمد. بنابراین متن باید به طریقی برای ماشین قابل خوانش شود. راحت‌ترین روش این است که معانی با اطلاعات همراه شوند.

****

سؤال اول:

این متن قانون چقدر توسط سیستم قابل فهم است؟

سؤال دوم:

چطور این فهم قابل توسعه است؟

برای این موضوع استانداردی به اسم آکوما انتوسو (Akoma Ntoso) را توسعه داده‌اند. با استفاده از قالب XML، استاندارد جدیدی را ساخته‌اند که بتواند اسناد پارلمانی، قانون‌گذاری، تقنینی و قضایی بین نهادها را پوشش دهد.

#### اهداف راهبردی:

1. ایجاد «زبان واحد» برای تبادل اسناد پارلمانی، تقنینی و قضایی بین نهادها
2. ذخیره‌سازی و دسترسی طولانی‌مدت به اسناد پارلمانی، قانونی و قضایی
3. ارائه خط‌مشی قابل اجرا برای سیستم‌های پارلمانی، قانونی و قضایی در مؤسسات
4. ایجاد الگوهای معمول داده و فراداده
5. ایجاد سازوکارهای مشترک برای نام‌گذاری و مرتبط‌کردن منابع
6. خودتبیین بودن
7. توسعه‌پذیر

پارلمان آمریکا درنظر داشته که آکوما انتوسو را توسعه دهد و متناسب با فرآیندها، ابعاد و نظام قانونی خودش متناسب‌سازی کند. کتابخانه کنگره ایالات متحده دو مسابقه برای تولید XML مبتنی بر Akoma Ntoso برگزار و برای یکی از آن‌ها 5 هزار دلار و برای دیگری 10 هزار دلار جایزه درنظر گرفت و اعلام کرد که هرکس بتواند محتوا را توسعه دهد، برنده خواهد بود.

# داده متصل (Linked Data)

اطلاعات چگونه به هم متصل می‌شوند؟

به عنوان مثال فردی از اطلاعات شهرداری استفاده و نرم‌افزار مسیریاب برای ویلچرداران طراحی کرد. خود شهرداری از ایده استفاده کرد و چیز دیگری ساخت. آن فرد از همین ظرفیت استفاده کرد و یک سایت برای مجموعه خیریه درست کرد. با پولی که بدست آورد دیتاست دیگری ساخت که آن هم با یک دیتاست در دانشگاه ترکیب شد و ... به همین ترتیب همه این‌ها به هم مرتبط می‌شوند به شرطی که دیتا این قابلیت را داشته باشد.

# کیفیت داده

همه داده‌ها یک‌جور نبوده و می‌توانند قابل استفاده باشند یا نباشند. برای این قابلیت، کیفیت انتشار داده را به پنج سطح دسته‌بندی کرده‌اند:

1. قالب‌های تصویری. صرفاً قابلیت خوانش انسانی میسر است یعنی داده یک ستاره‌. مانند فایل‌های PDF، JPEG، PNG
2. قالب‌های متنی. ویژگی این قالب‌ها این است که می‌توان از اطلاعات آن کپی‌برداری و همچنین در آن اطلاعات را جستجو کرد یعنی داده دو ستاره. مانند فایل‌های DOC، HTML
3. قالب‌های مشترک. قابلیت خوانش و پردازش اطلاعات بین انسان و ماشین یعنی داده سه ستاره. مانند فایل‌های XL، CSV، XML
4. فراخوان ماشینی. قابلیت فراخوانی و پردازش ماشینی یعنی داده چهار ستاره. مانند API (JSON)
5. رعایت استانداردهای داده متصل. تکمیل جورچین دادگان در ارتباط با داده‌های دیگر یعنی داده پنج ستاره. مانند API + Linked Data

در بیشتر کشورها حداقل سطح انتشار، سه ستاره و در بعضی دیگر 4 و 5 ستاره عنوان شده است. در ایران قانون و استانداردی در این زمینه وجود ندارد پس اطلاعاتی که داده می‌شود 1 یا 2 ستاره است و در موارد معدودی 3 ستاره می‌باشد.

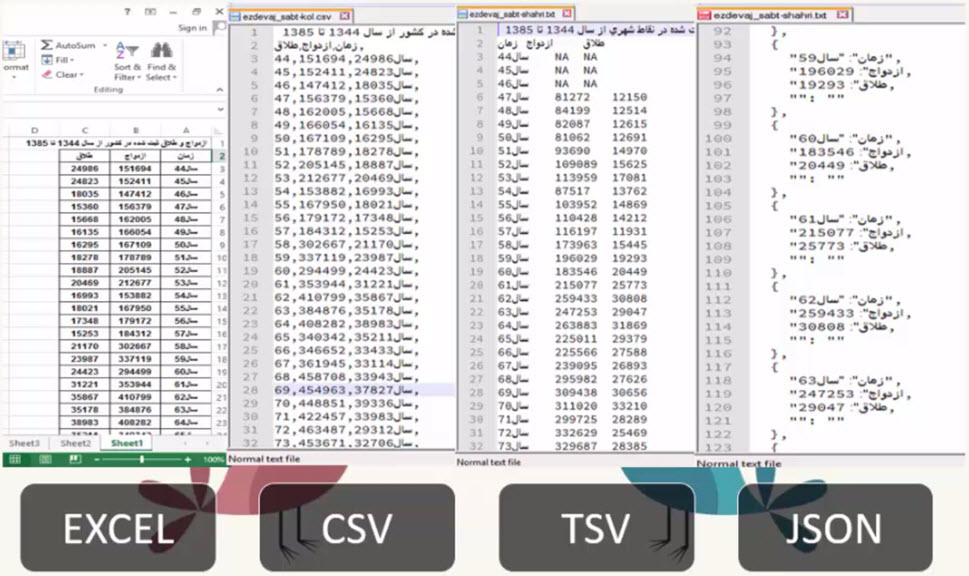
# مصورسازی داده Data Visualization

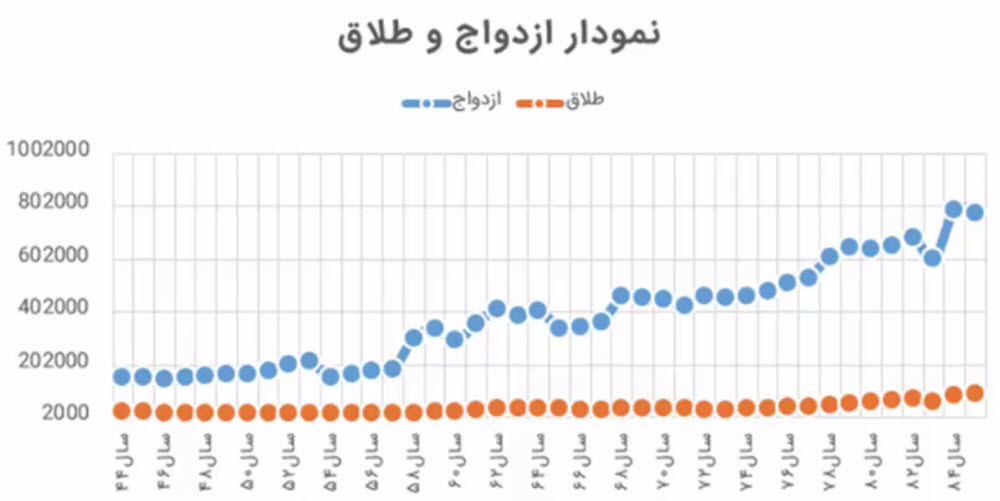
به‌طور خلاصه، مصورسازی داده یعنی اگر یکسری اطلاعات داده شود، در حالت عادی و در بهترین حالت تعدادی جدول پر از عدد هستند که نمی‌توان بررسی و تحلیل سریع و عمیقی از آن‌ها داشت.

## مثال

### میزان ازدواج و طلاق

در حالت عادی یک فایل اکسل است و همان اطلاعات به صورت CSV به شکلی دیگر است که با کاما از هم جدا شده‌اند. با فرمت TSV به‌جای کاما، تب دارد و یا در فرمت JSON، اطلاعات با API خوانده می‌شود. اما با مصورسازی این داده‌ها در قالب یک نمودار، به سادگی قابل فهم می‌شوند.





### توزیع بودجه فرهنگی کشور

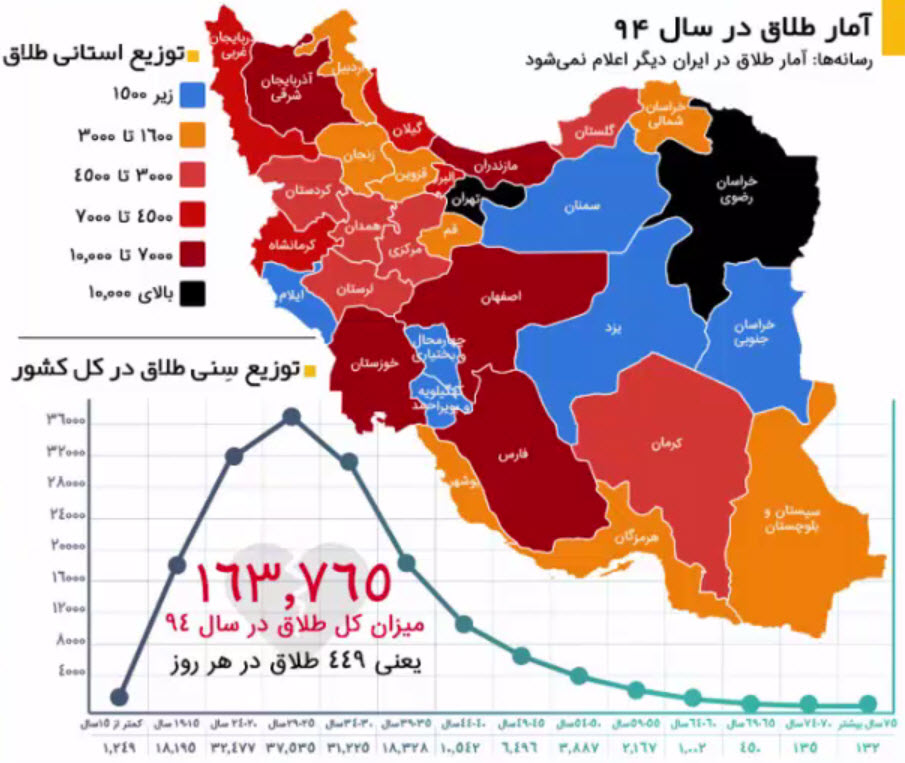
یکسری اطلاعات درباره بودجه‌های فرهنگی کشور را نشان می‌دهد که در سال چقدر بوده و چگونه توزیع شده است. اگر اطلاعات در قالب یک نمودار حجم نسبی ارائه شود تصورات دقیق‌تری را در پی خواهد داشت.





## نمودار ترکیبی

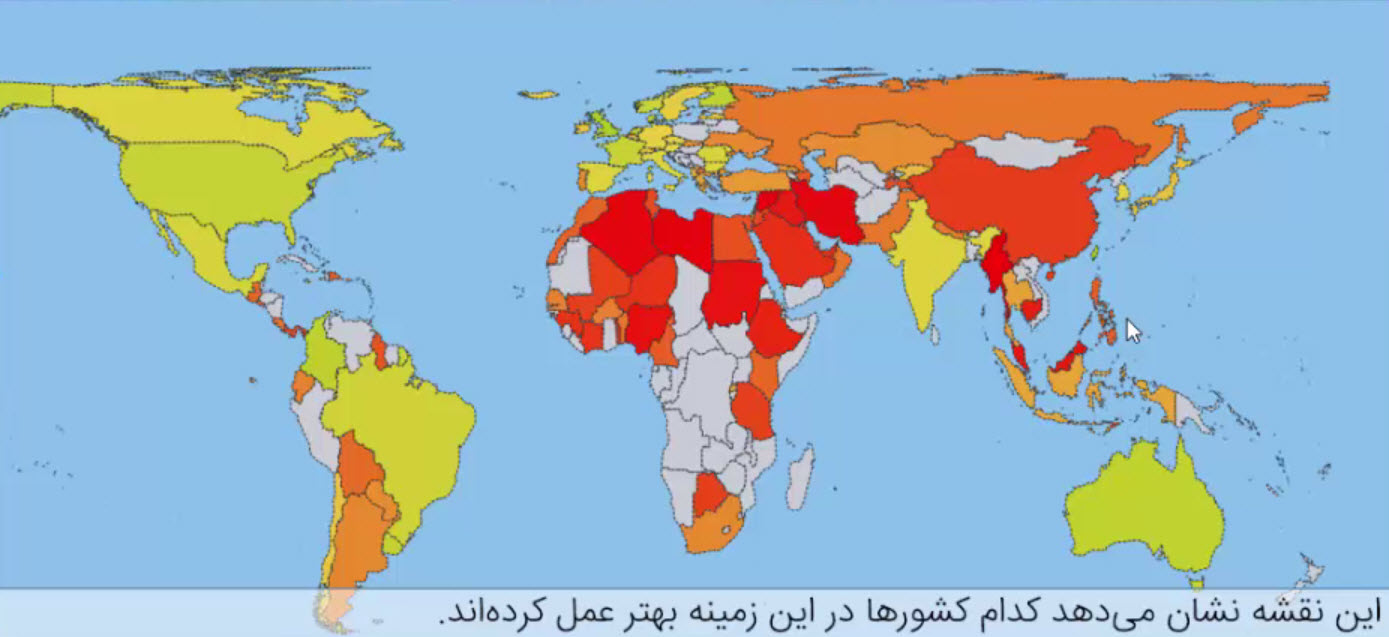
گاهی می‌توان نمودارها را با هم ترکیب و تصورات و اداراک‌های بهتری را ایجاد کرد. در نمودار زیر، آمار طلاق را با جغرافیا، رنگ و نمودار در کنار هم قرار داده و ادراک بهتری را ایجاد می‌کند.



# پورتال داده باز

آیا مطلوب است که از حاکمیت خواسته شود که در سایت‌های خود API قرار دهند؟ خیر. حاکمیت‌ها باید تلاش کنند که تمام داده‌ها در یک سایت، با یک استاندارد و الگو ارائه کنند.

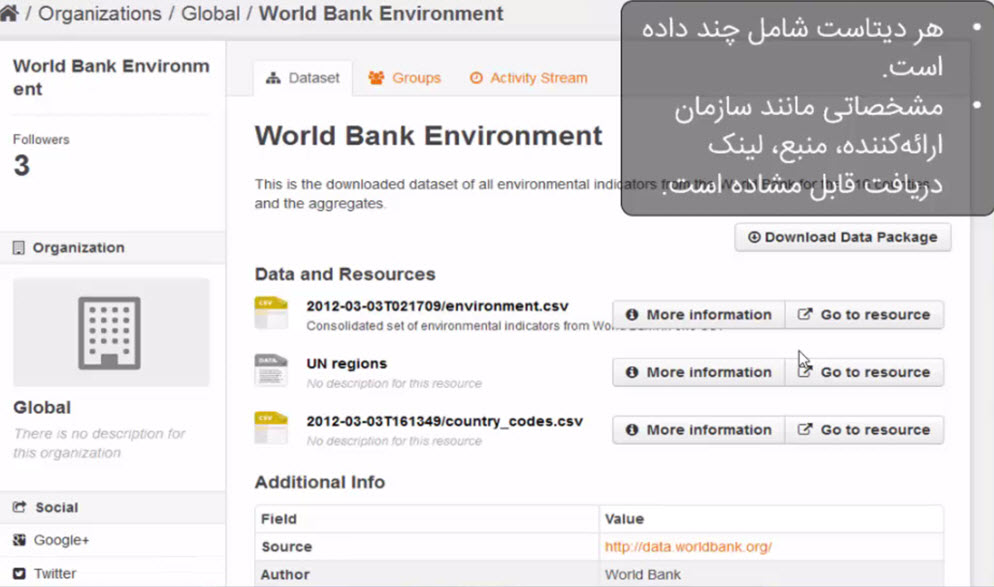
در نقشه وضعیت پورتال‌های داده باز کشورهای مختلف نشان داده شده است. ایران با رنگ قرمز مشخص شده و به این معنی نیست که اطلاعات را بد ارائه می‌کند بلکه به معنای این است که اصلاً چنین پورتالی را ندارد.



## چند مثال:

### پورتال datahub.com

یک نمونه سایت داده است که قالب‌های استانداردی مثل CSV را ارائه می‌کند. در این سایت همه اطلاعات یک قالب استاندارد دارند و هر دیتاست شامل چند داده است. همچنین مشخصاتی مانند سازمان ارائه‌کننده، منبع و لینک دریافت قابل مشاهده می‌باشد.



## پورتال data.gav

پورتال اطلاعات ایالات متحده آمریکا است که تمام داده‌های باز برای حکومت ایالات متحده در این پورتال قرار دارد. داده‌ها در موضوعات و سازمان منتشرکننده قابل دریافت هستند. قالب‌ها استاندارد می‌باشد. روش فنی مانند API وجود دارد و Linked Data های آن موجود است.

# «یو آر آی» (URI) چیست؟

عبارت است از یک رشته کاراکتر برای شناسایی یک نام یا یک منبع بر روی اینترنت.

URI از دو بخش تشکیل شده است:

1. URN: مشخص‌کننده نام منبع
2. URL: مشخص‌کننده روش دسترسی به منبع

