NuCivic Guide to DKAN

dkan

The all-in-one open data platform

nüCIVIC

Contents

DKAN

> "DKAN provides the optimal experience for a data management system: the ability to share your data in dynamic ways coupled with the means to tell stories and provide context that help bring it to life. Companies and government agencies around the world have understood that data alone, presented without demonstrating its potential, tells only a part of the story about how it was or could be able to create value for companies, governments, and individuals. DKAN excels at demonstrating that value."
>
> *-Rob Baker, former White House Presidential Innovation Fellow*

About DKAN

DKAN is an open source, open data platform with a full suite of data and content management cataloging, publishing and visualization features that empower organizations to easily distribute public data in easy-to-consume formats. It provides features that make it easy to upload, parse, store, publish, catalog and visualize a vast array of data, including spreadsheets, text documents and maps.

The DKAN project is maintained by NuCivic and developed openly and collaboratively with the open data, open source and open government communities. DKAN is licensed under the GNU Public License, version 2.

DKAN is powered by Drupal, an open source content management platform maintained and developed by a community of 630,000+ users and developers that includes 30,000 contributed modules and thousands of themes. It is aligned with the data standards and best practices of the CKAN data portal software and is fully compliant with leading open data standards such as DCAT and U.S. Project Open Data.

DKAN was identified by U.S. Project Open Data as a "ready-to-use" solution "that will help agencies jump-start their open efforts" and a "real, implementable, coded" solution "developed to significantly reduce the barrier to implementing open data."

In 2014, Amazon Web Services honored DKAN with its grand prize "City on a Cloud Partners in Innovation Award."

DKAN checklist

Data publishers

DKAN offers the following benefits and features for organizations publishing open data:

- Meets U.S. Project Open Data requirements
- Data.gov compability
- Provide metadata management capability, including automated generation of data.json files
- Full-scale data and content management system
- Full-featured visualization, graphing and mapping tools
- Secure, customizable data workflow management (roles, groups, permissions)
- Fine-grained user group, authorization and access controls
- Tiered access for public and internal audiences
- Easily customize the look and feel of the website with your own logo, color scheme and other branding requirements
- Customizable search functions
- Publish datasets in machine-readable formats using key open data standards including JSON, XML, RDF, DCAT and others
- Add individual datasets manually, and/or programmatically import data via API/harvesting from other applications
- Associate related files within each data set
- Federate data with other popular data platforms, including CKAN
- Store data either directly within DKAN or on external sites
- Automated harvesting, migration, importing of large-scale data
- Customize datasets by creating and managing your own metadata fields
- Section 508 Accessibility Compliant
- Unlimited licenses for users and managers
- Engage public end-users allowing them to comment, rate, share, embed and request a dataset
- Blend blogs, data, and other content to tell “data stories”
- Available on platform certified by FISMA, CSA, DIACAP, ISO 27001, and SSAE 16 SOC1 (SAS 70)

Data users

Researchers, journalists, entrepreneurs and developers can leverage DKAN in many ways:

- Easily download any data resource in multiple formats
- Preview spreadsheet data as tables, charts or maps
- Conduct comprehensive keyword searches on all data
- Access real-time APIs to efficiently develop web and mobile applications
- Rate, comment and give feedback on data resources
- Easily share data with others

DKAN features

Data management

DKAN is comprised of two powerful custom modules, DKAN Dataset and DKAN Datastore, which allow open data publishers to easily manage their open data publication process while complying with industry- and government-mandated standards.

Datasets

DKAN enables authorized users to create *datasets* which serve as collections of data *resources.*

Resources

A *resource* is an uploaded or externally-referenced file or API within a dataset. Resources can be either an uploaded file, the URL to a file that resides outside of the DKAN website, or the endpoint of an API that can be used to retrieve data programmatically from an outside web service.

By default, DKAN allows uploads of any file types desired. PDF or MS Word documents, HTML and image files can be treated as data resources if you so choose. You can also configure DKAN so that file uploads are restricted to only the file types that you choose to allow.

Datastore

Many of the data *resources* that are added to DKAN contain structured data. Spreadsheets, for example, organize data into individual data elements (often called *cells*), that are grouped into rows and columns. DKAN can extract this information and store it within a *datastore* that makes it possible for someone to search, select and, sort the individual rows, columns and data elements that make up the data resource.

The DKAN Datastore makes the data searchable, sortable and available for visualization in a variety of ways. Currently DKAN is able to parse the following file types (additional file types are in the process of being added):

- .xls
- .xml
- .csv
- GeoJSON

DKAN's datastore also allows users to export data as a comma-separated values (CSV) file that can be opened by any spreadsheet software. DKAN also provides an API that allows other software applications to retrieve data from the datastore.

Open data standards compliance

DKAN provides a "data.json" index to satisfy the U.S. federal government's Project Open Data requirements. More information about the "slash data" or "data.json" requirements can be found in POD's Open Data Catalog Requirements and Common Core Metadata Schema pages.

The exact mapping of data from your DKAN site to the data.json index can be customized using DKAN's Open Data Schema Mapper module, which allows for easily including metadata fields in public read APIs. The Open Data Schema Mapper comes with built-in schemas for CKAN and Project Open Data and provides an interface through which other, custom schemas can be easily added.

User roles and authentication

DKAN includes a sophisticated user roles and permissions system that provides a full graphical user interface for creating and managing user accounts, user roles, and workflows. For user authentication, administrators have the option of using DKAN's native one or two-factor authentication, which optionally can support oAuth-based social/distributed login. For staff authentication and/or single sign-on, DKAN can be customized to integrate existing internal authentication systems, such as Lightweight Directory Access Protocol (LDAP) or Security Assertion Markup Language (SAML).

Built-in DKAN user roles include:

- Portal administrator
- Group organizer
- Content editor
- Data contributor

Additional user roles and permissions can be created as needed.

Data publishing workflow

DKAN's Data Workflow modules provide a combination of roles, permissions, content views and notification rules that provide a basic editorial workflow to data publishers. These roles and permissions can be enabled or disabled independently of one another depending on which are actually needed.

DKAN's workflows are built using the Organic Groups module for Drupal, which means that they can be easily customized. Roles and permissions can be tailored to your organization, its editorial policies, and its desired workflow. For example, you may want to set permissions so that only editorial staff an view and edit datasets until they have been approved for publication. You can also create private groups whose contents are only visible to group members.

Design customization

DKAN ships with an attractive responsive design, but you are not restricted to using the built-in design. Your DKAN open data platform can be easily customized by adding your own logo and color scheme or by inserting your own CSS and adding custom page layouts. For full control over the look and feel of your website, you can create your own theme entirely or adapt one of the thousands of themes that are available as free downloads from the Drupal website.

Content management

DKAN is developed using the popular Drupal software framework and leverages all of Drupal's inherent strengths as a content management system, such as page and blog creation, user roles and permissions, all within an intuitive user interface front-end and WYSIWYG editing.

Key DKAN content management features:

- full graphical user interface for administering content, workflows, roles and permissions
- blogs
- topic-based groups
- taxonomies (tagging, categorization)
- faceted search
- form-building
- calendars
- and more

Metadata

DKAN uses DCAT (Data Catalog Vocabulary), a standard vocabulary for describing datasets in data catalogues. DCAT is the basis for the metadata schema prescribed by Project Open Data. It provides RDF endpoints and RDFa markup for all Datasets, following current best practices designed to make data more discoverable and enabling other software applications to combine or search across multiple data catalogues. In addition, DKAN allows for custom metadata fields that can be added to datasets and resources as needed.

DKAN datasets come with the following default metadata field options that can be expanded based on your custom needs:

- author
- contact information
- data release date
- geographical area to which the data applies
- licensing information
- and more

Users can also add customized metadata fields to DKAN which will show up on the dataset form, in search facets, and available to output in one of the dataset APIs.

Preview

DKAN allows the public user to preview the data directly on the DKAN website. It can display in a spreadsheet format (as a scrollable table that is searchable and sortable). Quantitative data can be displayed as bar charts, line charts and other types of graphs. If the data includes geodata such as latitudes and longitudes, the data can also be displayed as markers on a map.

Search

DKAN offers rich faceted search using Drupal's Search API module, which makes it possible to refine and narrow search results based on keywords, data type, location, author, and other categorization and classification criteria.

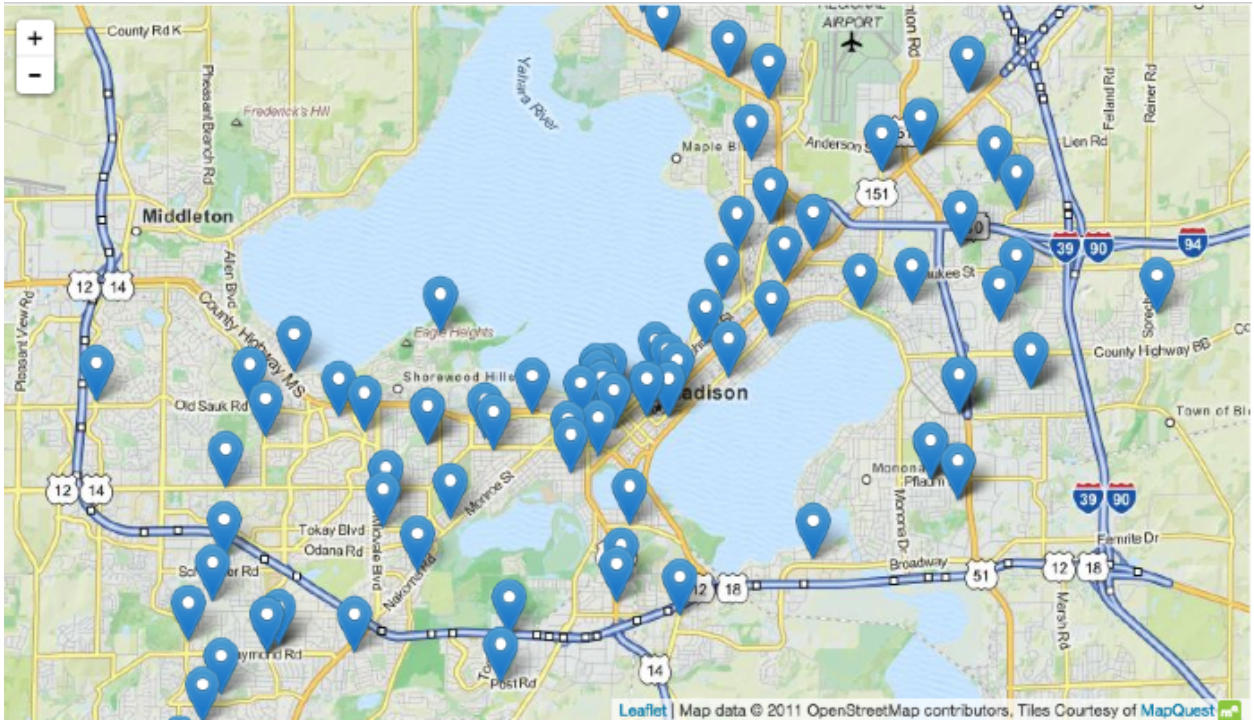
By default, DKAN uses Drupal's native database to index content for searching. For more powerful search, DKAN can be easily configured to use the Solr search engine using the Search API Solr module.

Migrate data

To facilitate importing of data from other sources, DKAN provides a "harvester" module that leverages the Drupal Migrate module. Data imports can be performed on an as-needed basis or can be set up to automatically refresh from data feeds at regular intervals. With a small amount of custom setup work, data can be harvested from a wide range of existing data sources and data formats.

DKAN's Dataset API, which leverages the Drupal Services module, can also be enabled to provide full CRUD ("create, read, update and delete") support for Datasets and Resources. In addition to importing data into DKAN, this enables other software applications to retrieve data from DKAN. The result is that your data is truly free. Wherever you choose to store your data, DKAN can retrieve, catalog, share and display it.

Visualization



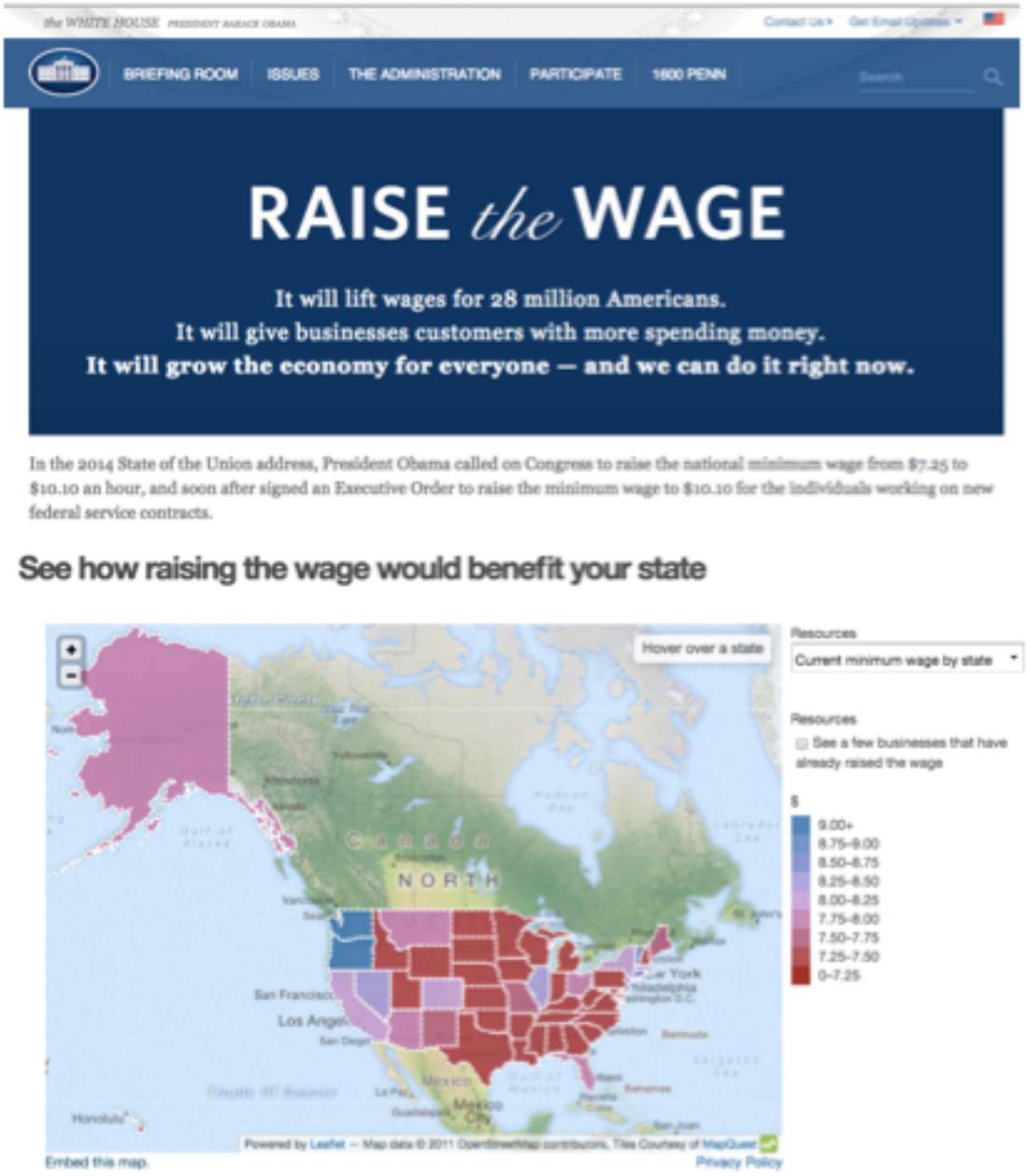
An example of a DKAN map visualization showing election polling stations in Madison, Wisconsin.

The DKAN data visualization entity incorporates the Recline.js library and choropleth mapping into its core functionality, providing intuitive graphing and mapping capabilities to users.

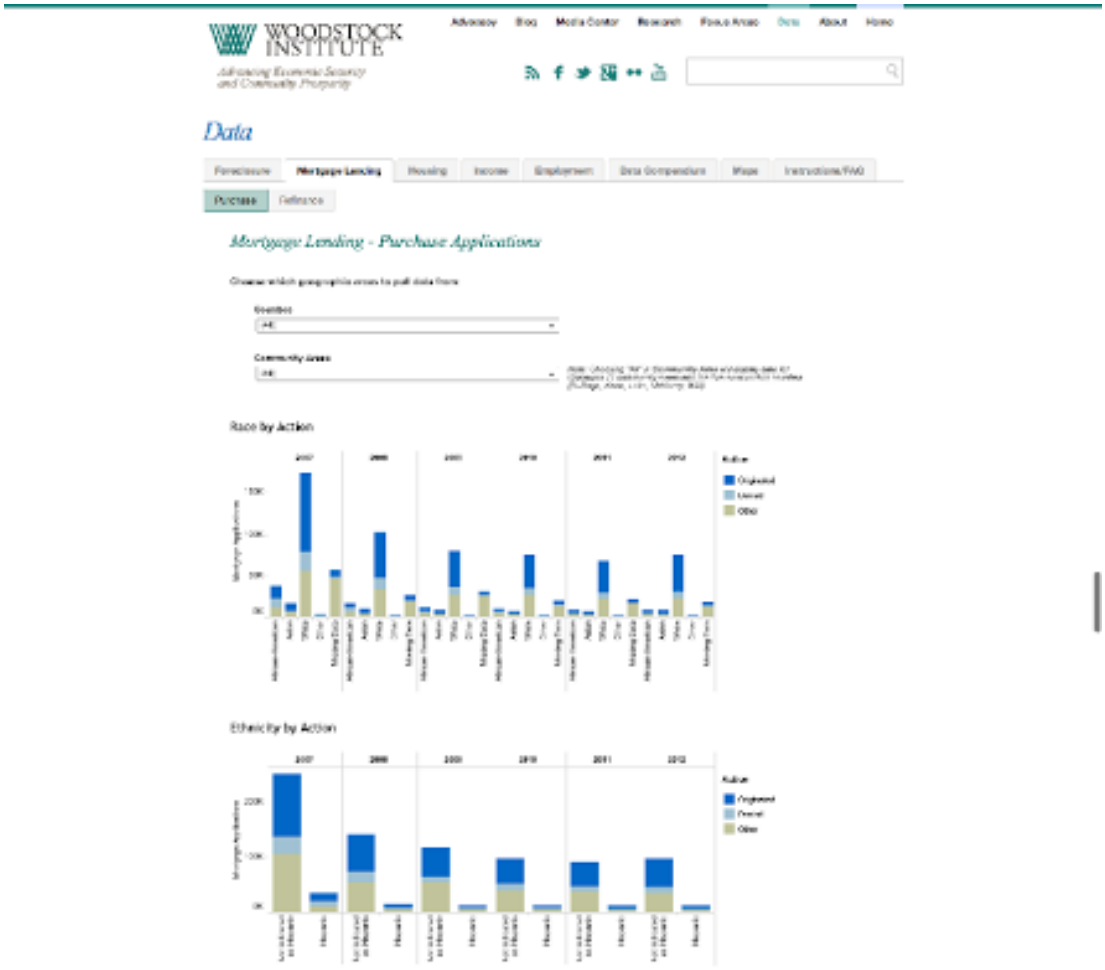
Recline.js is a powerful data visualization library that provides tabular, chart, and (for spatial data) map previews of any data uploaded into DKAN. Recline.js also provides end-users with various tools to search and filter data, as well as to create widgets that let you share, download and embed data visualizations in other web pages.

DKAN can also be custom integrated with a wide range of visualization software programs and tools such as:

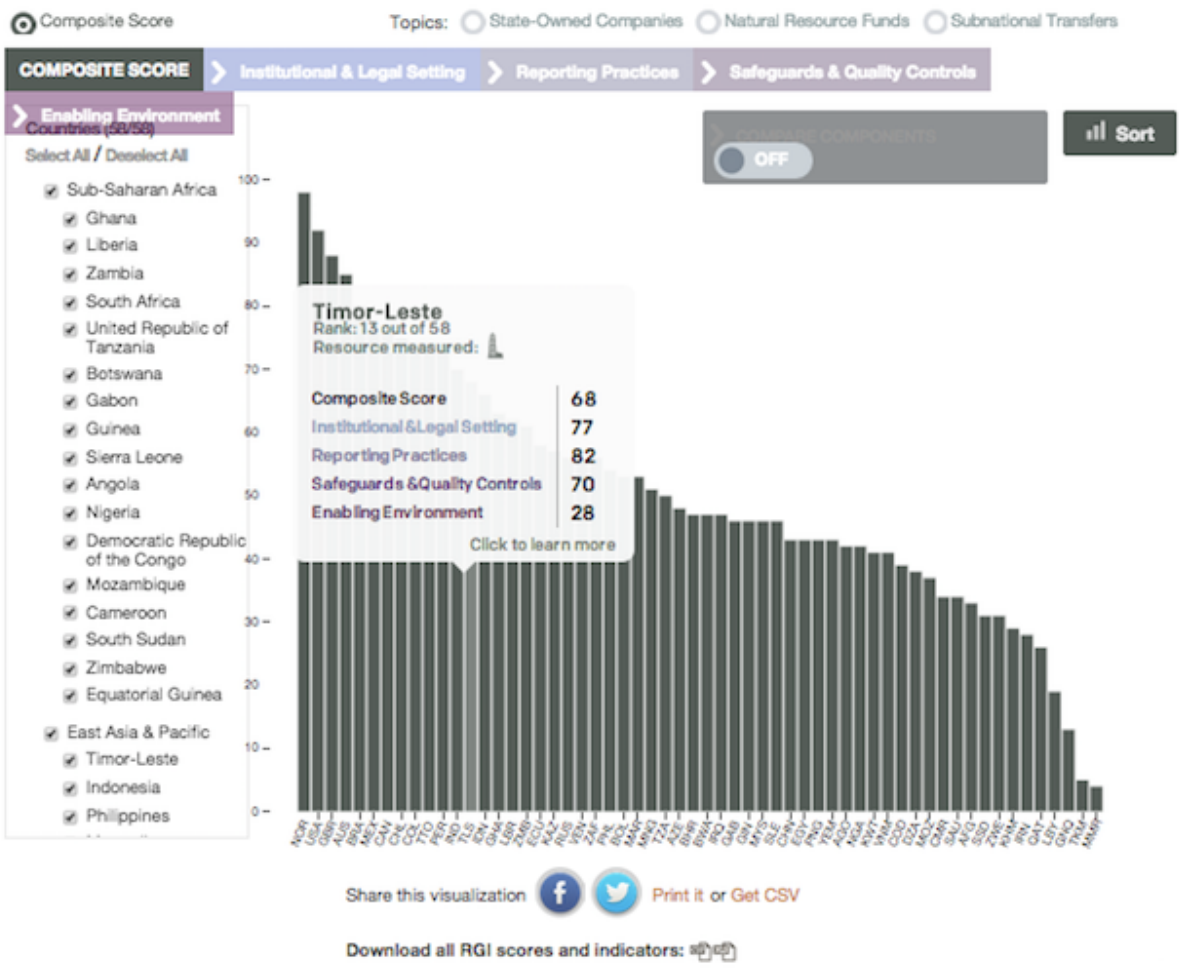
- Tableau
- CartoDB
- Crystal Reports
- D3
- and more



The White House used this DKAN-generated choropleth map to support its "Raise the Wage" campaign which made the case for raising the minimum wage in the United States.



The Woodstock Institute, a NuCivic client, uses Tableau, a popular data visualization software package, to create interactive graphics like this one and embed them within individual web pages.



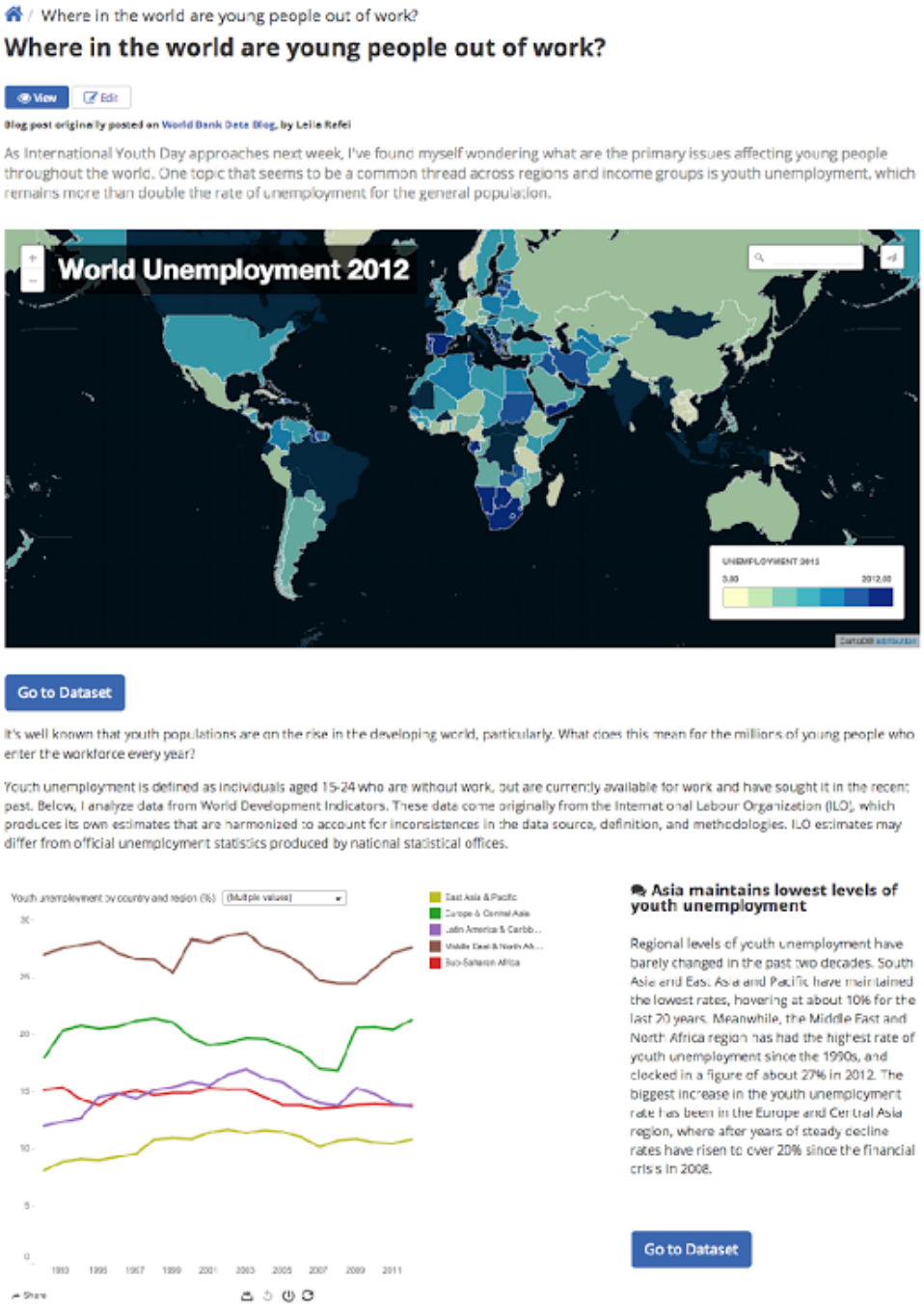
NuCivic used D3.js javascript technology integrated with the DKAN Datastore, to provide rich interactive data visualizations customized for the Resource Governance Institute's specialized extractive industry data.

Data stories

Telling stories with your data using photos, videos, maps, charts, graphs and visualizations is easy to do leveraging the DKAN content management system and embed tools that integrate and link directly to the referenced data.

This makes it easy for data managers to give better context to their data and support those stories and showcase the power of open data. You can tell the story and illustrate it using all the tools of DKAN's rich, Drupal-based content management system while providing links that let power users connect to and retrieve the raw data so that they can do their own analyses and find new uses for the data.

Data stories are a great way to put data in a more accessible intuitive context for many users, provide strong SEO for search engines, and still maintain the key links back to the underlying data.



This example data story helps visualize and unemployment patterns with text, a map and a comparative time chart, combined with "Go to Dataset" links that make it easy to dig deeper into the underlying data behind the story.

Extending DKAN

DKAN can also be easily extended through Drupal by adding some of the 30,000 modules that are available as free downloads from the Drupal website. Combining DKAN's built-in functionality with this rich collection of optional add-ons gives you an agile, flexible and cost-effective open data platform.

DKAN technical specifications

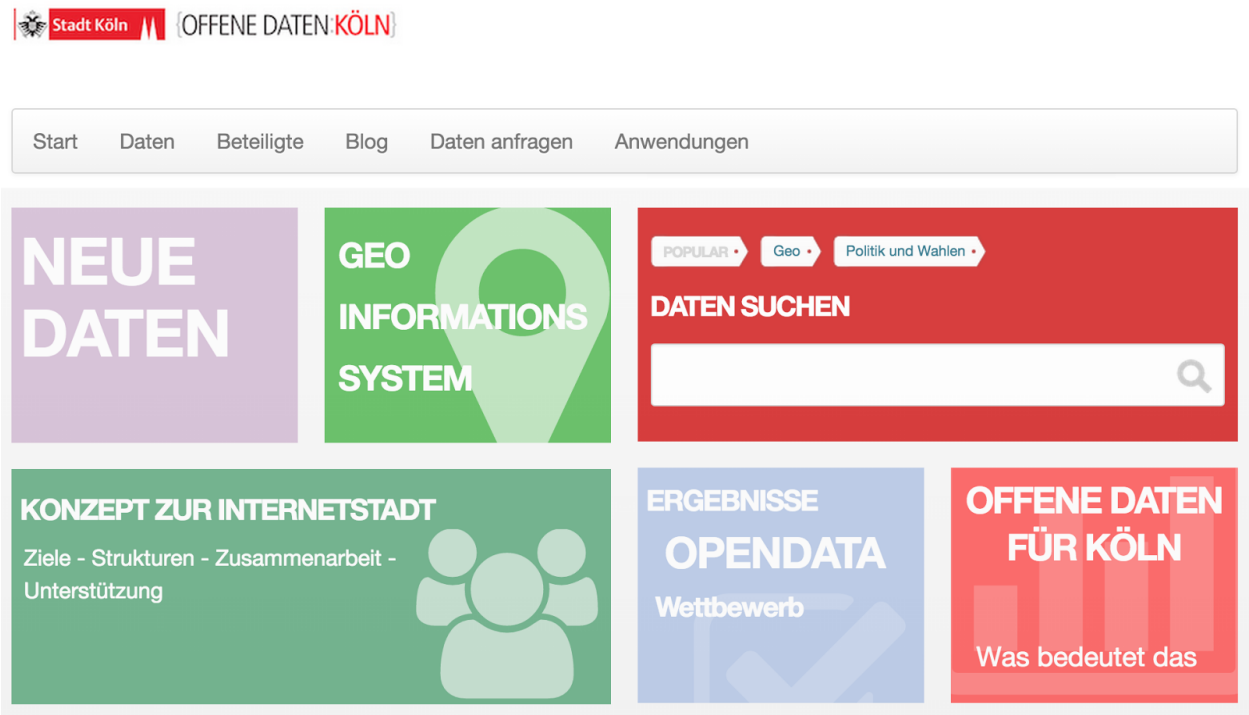
DKAN is developed in the Drupal content management platform, which is powered by the LAMP stack (Linux, Apache, MySQL and PHP). However, it is flexible enough to run on other operating systems and to use other web servers and database backends if desired.

Technical specifications:

- Platform: Drupal
- Primary language: PHP
- Database: MySQL, MariaDB, PostgreSQL, SQL Server, or Oracle
- Web server: Apache or Nginx
- Operating system: Linux, Windows, OSX, or Unix

DKAN case studies

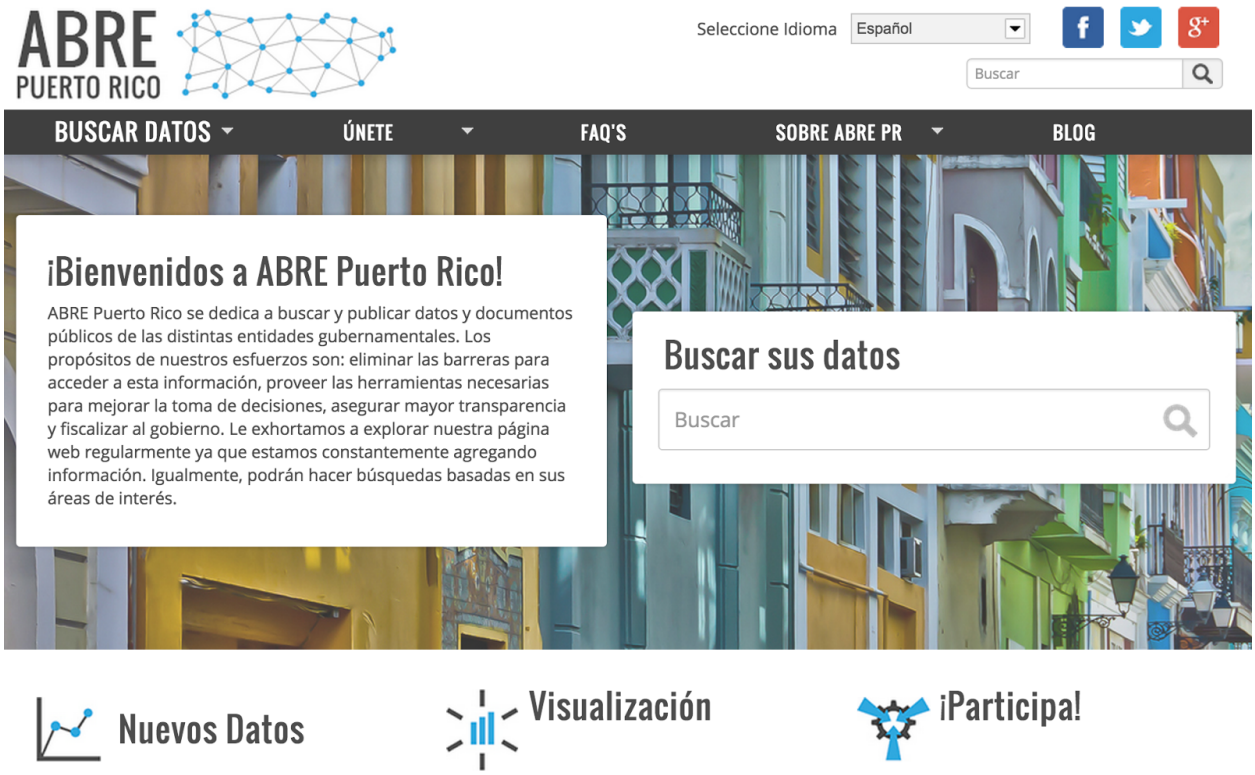
City of Cologne (Germany)



NuCivic provided technical support and customization engineering to the technology team of Cologne, Germany, for their DKAN open data portal. In addition to training their technical team, we have assisted them in customizing their Solr search, dataset schema and API to better meet their data management requirements in a manner that aligns with DKAN best practices and its roadmap. Finally, we assisted Cologne in federating the data in its data portal up to the national German open data catalog, which is based on the related CKAN platform.

Site: http://www.offenedaten-koeln.de

Open Puerto Rico



NuCivic designed and built a DKAN-based multilingual open data portal for the Abre Puerto Rico non-profit open government organization, helping to bring a new level of transparency and accountability to the government of Puerto Rico. A particular challenge on this project was the full multi-lingual (Spanish/English) requirement, where we leveraged Drupal's internationalization modules (https://www.drupal.org/project/i18n) to deliver a seamless experience for users in either language.

Site: http://www.abrepr.org

DKAN resources

- DKAN demo site
- DKAN documentation
- DKAN developers group
- DKAN on GitHub

Glossary

Choropleth: A thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income. *(Source: Wikipedia)*

Dataset: A collection of data resources (see Resource).

DKAN: Open data platform built on the Drupal open source content management system.

ETL: Short for "Extract, Transform and Load," allows data managers to automate data publishing from enterprise systems to an open data platform.

NuCivic Data: Cloud-based, fully supported offering of the DKAN open data platform, managed by NuCivic.

Recline.js: A powerful library for building data applications, used in both DKAN and CKAN to quickly preview and visualize datasets in the browser.

Resource: An uploaded or externally-referenced file or API.

NuCivic

About NuCivic

NuCivic is a civic-focused software-as-a-service company that provides turnkey open source solutions to public sector institutions around the world.

NuCivic was co-founded by Andrew Hoppin and Sheldon Rampton in 2011, inspired by their work introducing and deploying open source systems for the New York State Senate, where Hoppin served as chief information officer. The NuCivic team is comprised of experienced open source, open data, cloud operations, and open government professionals that bring strategic, philosophical and technical expertise from all areas of the civic spectrum.

Federal Computer Week named NuCivic one of "17 hot companies to watch."

In December 2014, NuCivic was acquired by GovDelivery Inc., to help scale the business and meet growing local, state, and federal government demand for cloud-based, open source solutions. GovDelivery provides an enterprise-class, cloud-based platform that allows more than 1000 government clients to create and send billions of messages to more than 70 million people around the world.

Contact

- Phone: (917) 426-9136
- Email: sales@nucivic.com
- Web: http://nucivic.com

Connect

- Blog: http://nucivic.com/blog
- Subscribe: http://nucivic.com/subscribe
- Twitter: @getNuCivic
- GitHub: http://github.com/nucivic