

# Our Next Challenge: Open Data Platforms That Talk to Each Other

## Chris Metcalf

Director of Product Development and Developer Evangelism  
Socrata, Inc.  
Seattle, WA

## Synopsis

Some of the biggest challenges the field of open data faces as it matures are those of discoverability, interoperability, and federation. In this talk, Chris Metcalf, Director of Product Development for leading cloud-based open data platform provider Socrata, will describe some proposed methods by which we as open data practitioners and implementers can address them.

## Challenges in Open Data

As open data becomes more a part of how governments, nonprofits, and NGOs function, the following issues must be addressed:

- **Discoverability** - In order for data to be useful, it first needs to be discoverable. Building strong catalogs of metadata from numerous sources is one of the best ways this can be achieved. In addition, building in referenceability into data catalogs is also an effective way of tracking how data propagates through different work products - from a raw dataset to an API to a mashup, etc.
- **Federation** - As open data becomes part of the day-to-day business of these organizations, the work of cultivating, publishing, and maintaining datasets and data catalogs will become more decentralized. These decentralized catalogs still need to be aggregated into combined organizational catalogs, but their maintainers should still be able to pick and choose which open data technologies are most appropriate to their needs. In time, these data catalogs will stop resembling the big, centralized catalogs being built now, and will morph into an ecosystem of distributed, federated data catalogs that need an effective and efficient way for catalogs to be propagated for discovery.
- **Interoperability** - As catalogs from multiple sources are composed by federation, it becomes more and more important for the platforms that these data catalogs are built on to be compatible, even if they are built by different providers.

## Developing Cross-Platform Dataset Metadata Standards

Open data platform providers like CKAN, Junar, and Socrata must work together to build cross-platform data catalog standards that are implemented consistently between different catalog platforms. Developing practical, portable data catalog standards will:

- Allow data catalogs to be easily federated and shared between organizations. This will, for example, allow a parent organization to easily consume and aggregate catalogs from different sub-organizations, regardless of what software those catalogs are hosted on.
- Allow catalogs to be portable between providers. As organizations change and grow, they may also want to move from one provider to another. Standardized metadata schemas make that a more straightforward process.
- Allow the catalog to track reuse and propagation of data sets through the catalog. By allowing catalog entries to be related to one another, the catalog will morph from a simple listing of data sets into a network of related entries, supporting reuse and discovery.
- Allow data catalogs to be consistently indexed by search providers. Google and other search engines will be able to consume structured data catalogs and provide better insight and discovery into what data sets are available.

Work is already under way on efforts to build these standards. Through projects like DCAT, Schema.org, and the Open Data Substrate, we're making headway towards making the hard decisions necessary to build out interoperability standards.

But let's not allow ourselves to believe that simply having standards is enough. The world doesn't need more unimplemented standards. The best standard is the one with actual running, production-level code. Let's use this opportunity together to hammer out the details necessary to solve the real-world problems we face, and hopefully walk away with a more connected open data ecosystem.