Beyond Websites:

Drupal as Data Pipeline for Digital Signage

Mike Madison

About Me

Technical Architect @ Acquia
Organizer for Drupal GovCon

D.O: mikemadison
Github: mikemadison13
LinkedIn: mikemadison


AUGUST 22-24 • BETHESDA
FREE ADMISSION
OPEN TO ALL • PUBLIC & PRIVATE
DRUPAL
GOVCON
2018
GOVCON 17
DRUPAL
#DRUPALGOVCON
DRUPALGOVCON.ORG

Today's Agenda

- Methodologies for building a data pipeline
- Testing with Behat / PHPunit
- Integrating Drupal and its features / capabilities
- Powering Digital Experiences (broadly)

Today's Talk

- Advanced Topic… Intermediate Level
- Architectural Discussion

Disclaimer: Technology

Image Source: http://www.nydailynews.com/

BUZZWORDS

- Decoupled
- Headless
- Digital Experience

Why do you need a Pipeline?

"To create a persistent digital experience, you need consistent data."

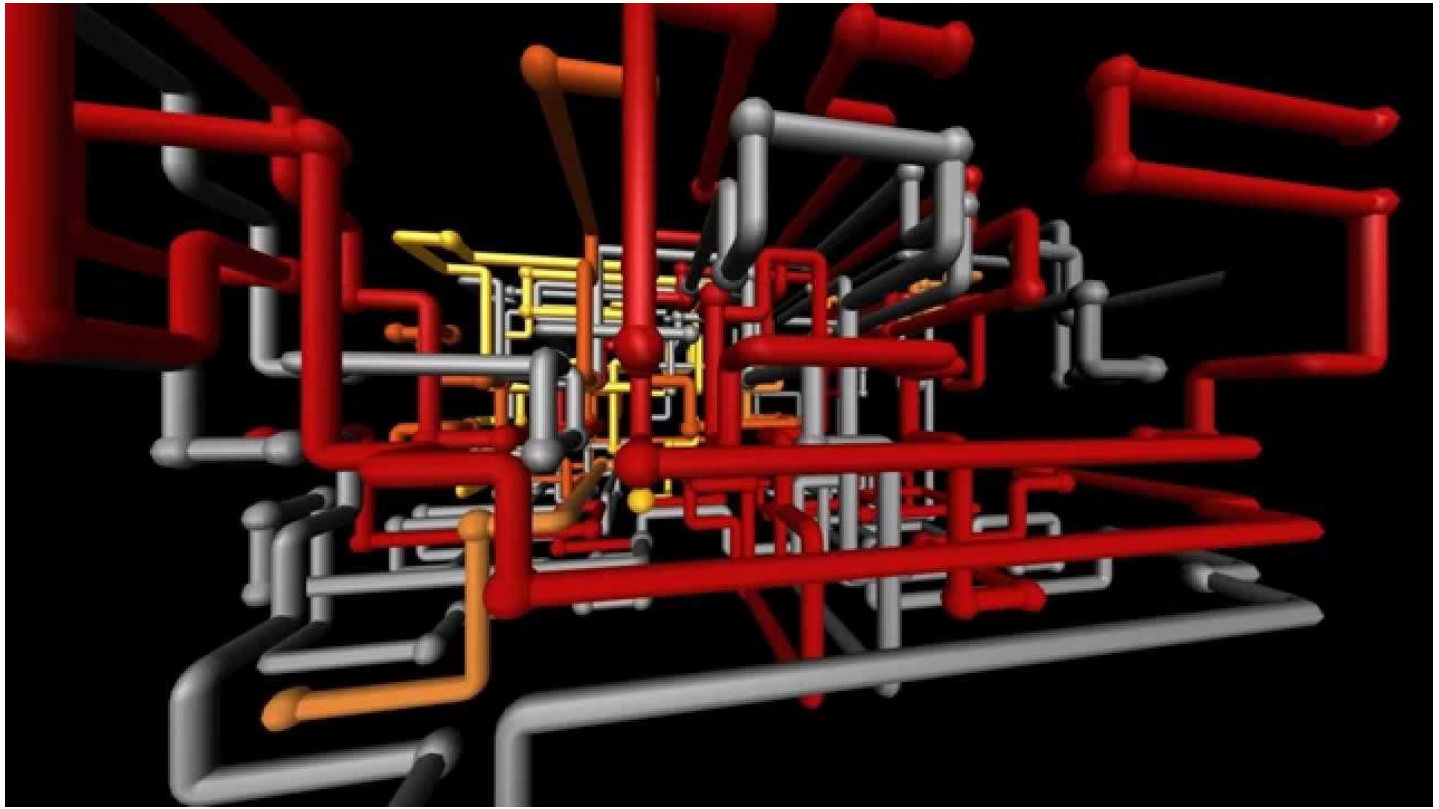
What Does a Pipeline Do?

- Pushes Data Automatically
- Limits Human Interaction
- Streamlines Data Collection
- Broadcasts to Multiple Location

What Does a Pipeline Do?

- Limits what/who can interact with system
- Provides fault tolerance

Methodologies for Building a Data Pipeline



Methodologies for Building a Data Pipeline

1. What is coming into the pipeline?
2. What is the pipeline powering?
3. How is the <whatever> displaying the <whatever>?

Data

Details

- What is it?
- How often does it update?
- Are there failover URLs?
- What format is it in?
- Do you have to authenticate to get to it?
- How big is the data dump?
- Are there API limitations / requirements?

Example

1. What is the data -- Mass Transit Data
2. How often does it update -- Every few seconds
3. Are there failover URLs -- Yes (multiple)
4. What format is it in -- JSON
5. Do you have to authenticate to get to it -- Yes
6. How big is the data dump - <10mb

Example Impact

1. ## Mass Transit Data
   NOT something you can screw up (**people might die**)

2. ## Every few seconds, <10mb, JSON
   FAST processing time in a standard format (small)

3. ## Multiple URLS, Authenticated Requests
   Secure + Robust

Digital Experiences

Details

- What is the Digital Experience?
- Can people interact with it?
- Is there any personalization / contextualization?
- How often does it need to update?
- Is it language dependent?
- Is it ADA compliant?

Example

- What is the Digital Experience? Arrival Sign
- Can people interact with it? No
- Is there any personalization / contextualization? Contextualization based on Physical Location
- How often does it need to update? Every 5-8 secs
- Is it language dependent? No
- Is it ADA compliant? Yes

Impact

- Arrival Sign, Every 5-8 secs
  Old data not useful (real world evidence if right / wrong)

- Contextualization based on Physical Location
  Sign must display appropriate location data

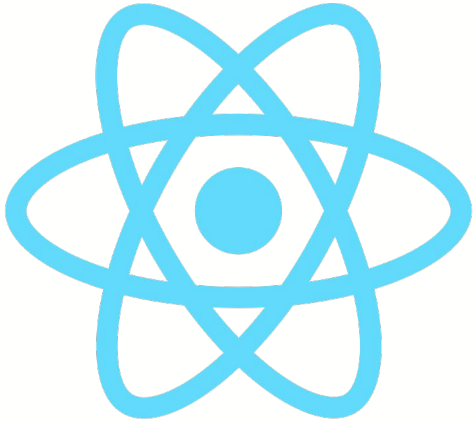
- No Interaction
  Sign has to just “work”

Impact

- Not language dependent
  BUT IT COULD BE

- ADA Compliant
  Heavily limits how much “stuff” can be on the screen at any one time

Front End



Cool! Now what?

Architecting a Solution

- Pipeline:
  - Acquire Data
  - Organize Data
  - Send data

Architecting a Solution

- Context:
  - Structure for organizing???

Architecting a Solution

- Communication with Devices:
  - How do you keep devices connected?
  - How do you get data from the pipeline to the device?

Pipeline: Data Acquisition

- Authenticate
- Get Data From <source>
  - Get Data from <other source>
- Validate Data

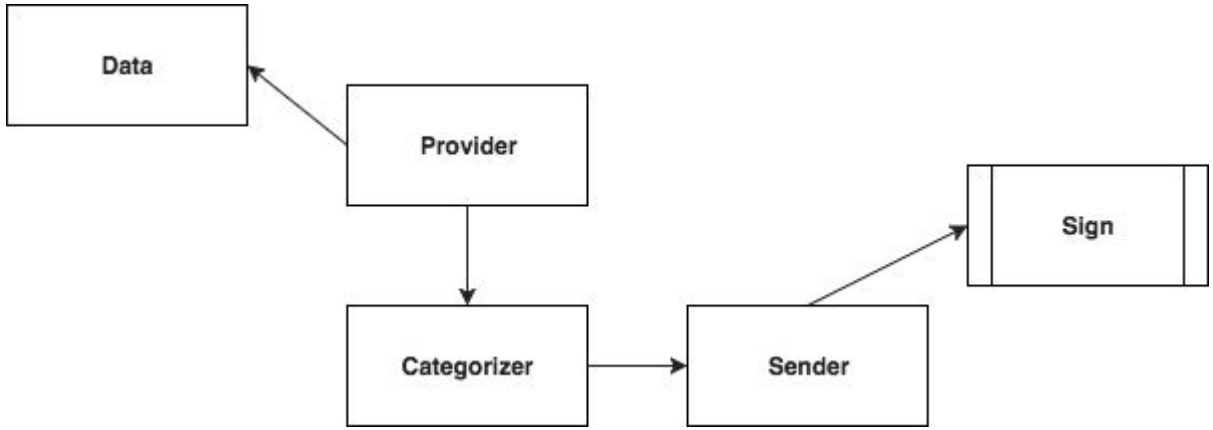
Pipeline: Categorize Data

- Isolate Data for Each <whatever>
- Break Up Data Into Smaller Payloads

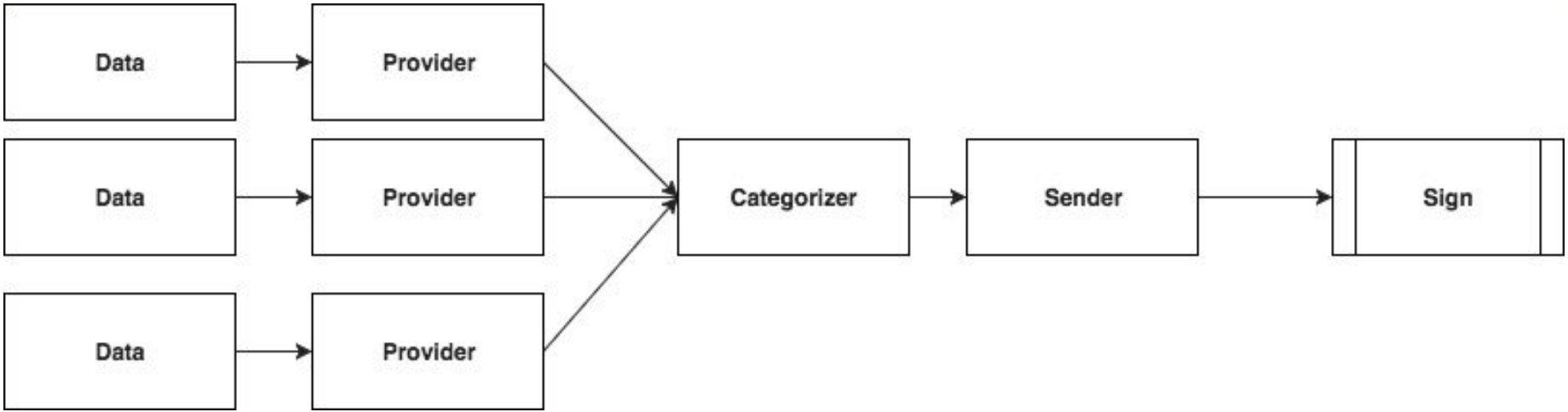
Pipeline: Sender

- Authenticate Into Appropriate Service
- Send the Data

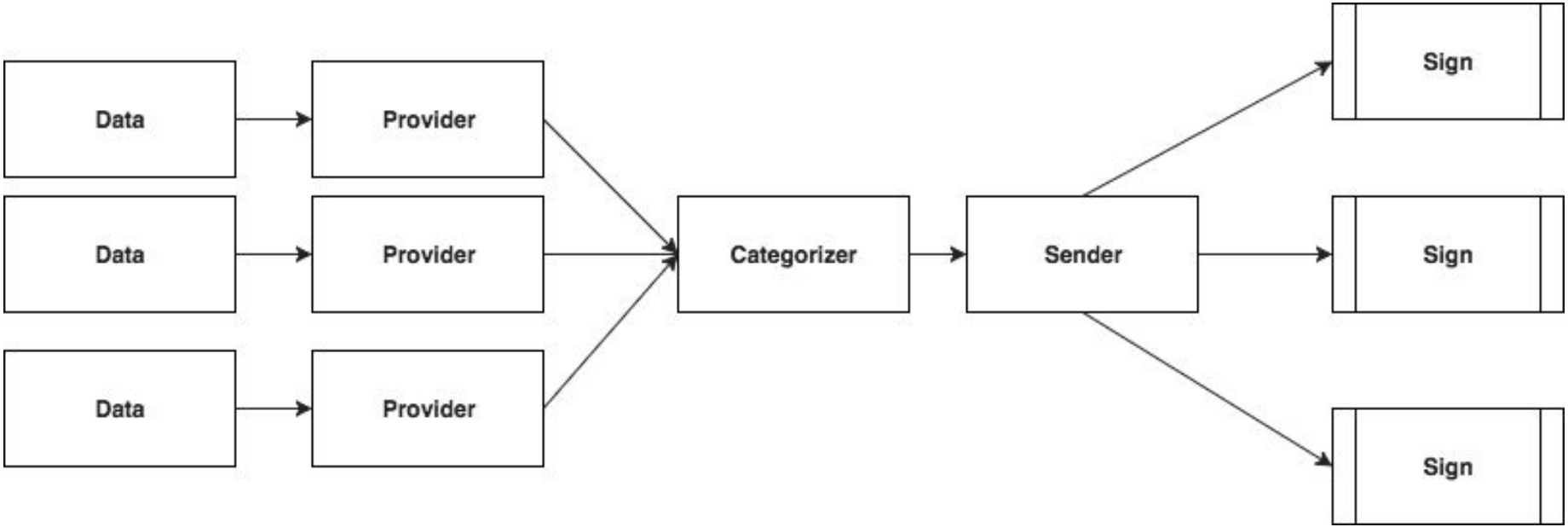
Pipeline



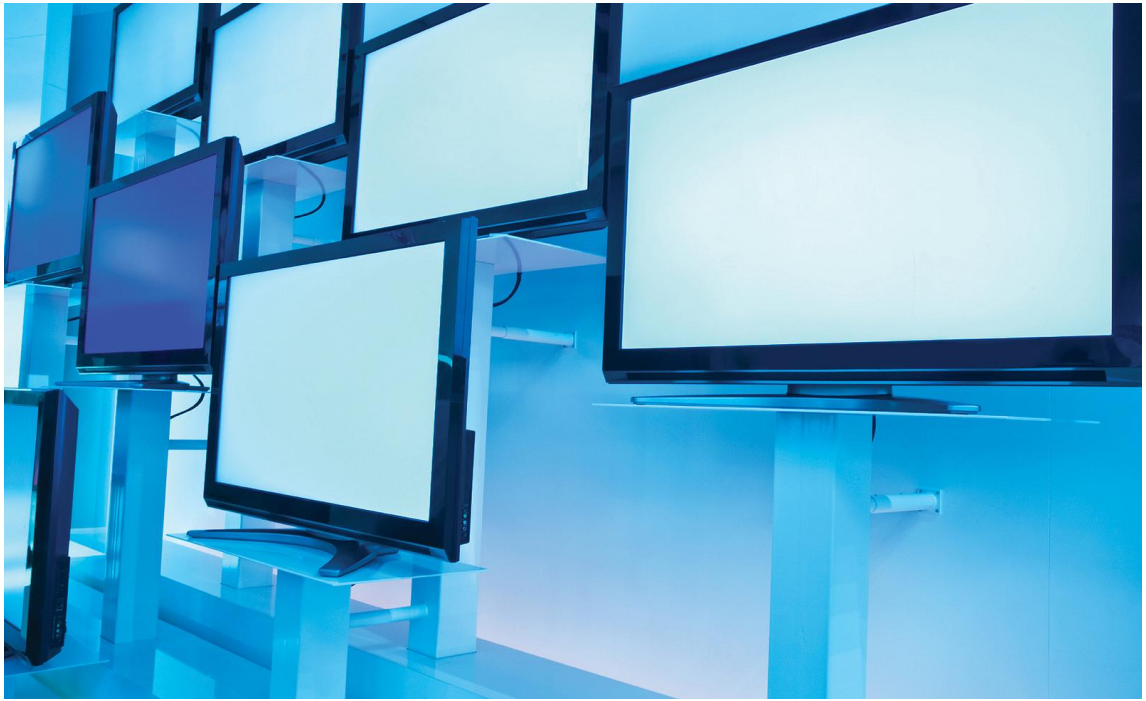
Pipeline



Pipeline



Lots of Signs Means...



https://www.flickr.com/photos/83428770@N06/

Lots of Little Details

- Size / Orientation
- Physical Location
- Facing Direction
- IP Address

OK Fine, Let's Use Drupal

- Drupal can manage this structure very easily

Remember...

- Workflow
- Permissions / Roles
- WYSIWYG / Easy to Use Forms
- Contrib
- Security
- Media
- Views / Reporting

Drupal as a Framework

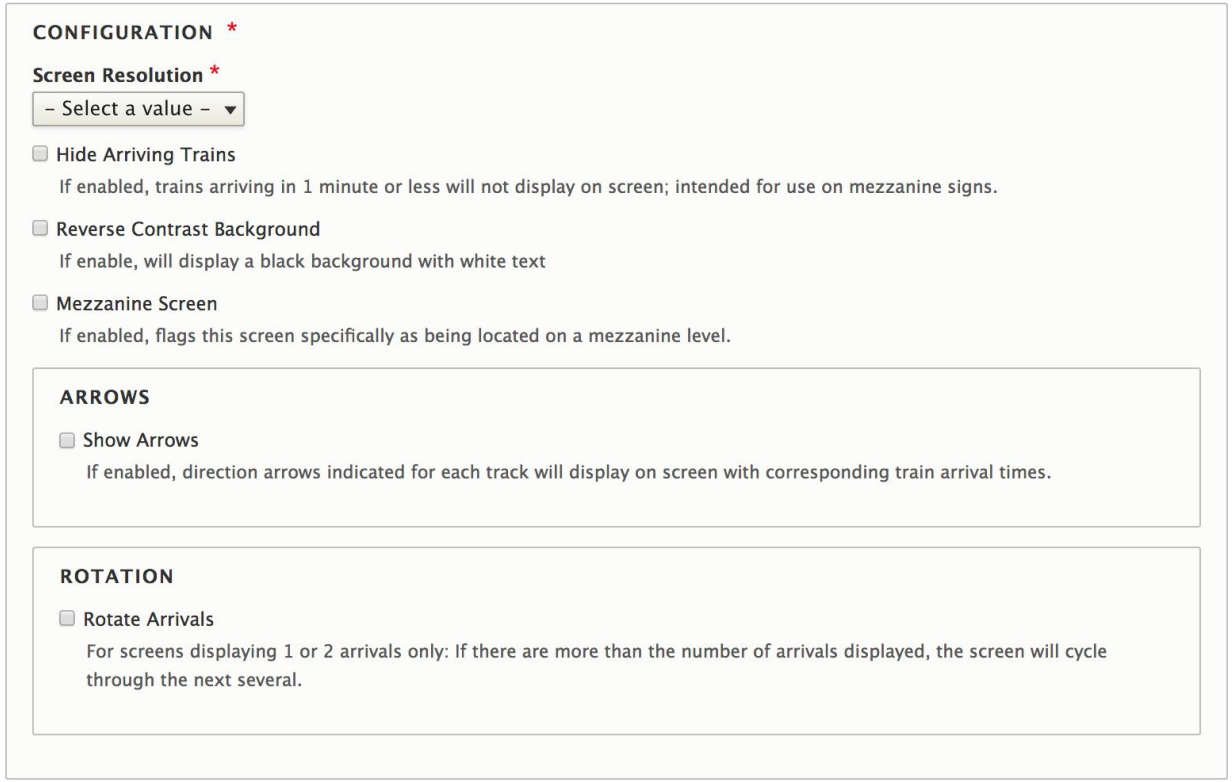
Content Types to Define Relationships

- Route Groups
  - Routes
    - Stations
      - Platforms
        - Signs

Content Types to Define Relationships

- Route Groups
  - Routes
    - **Stations**
      - **Platforms**
        - Signs

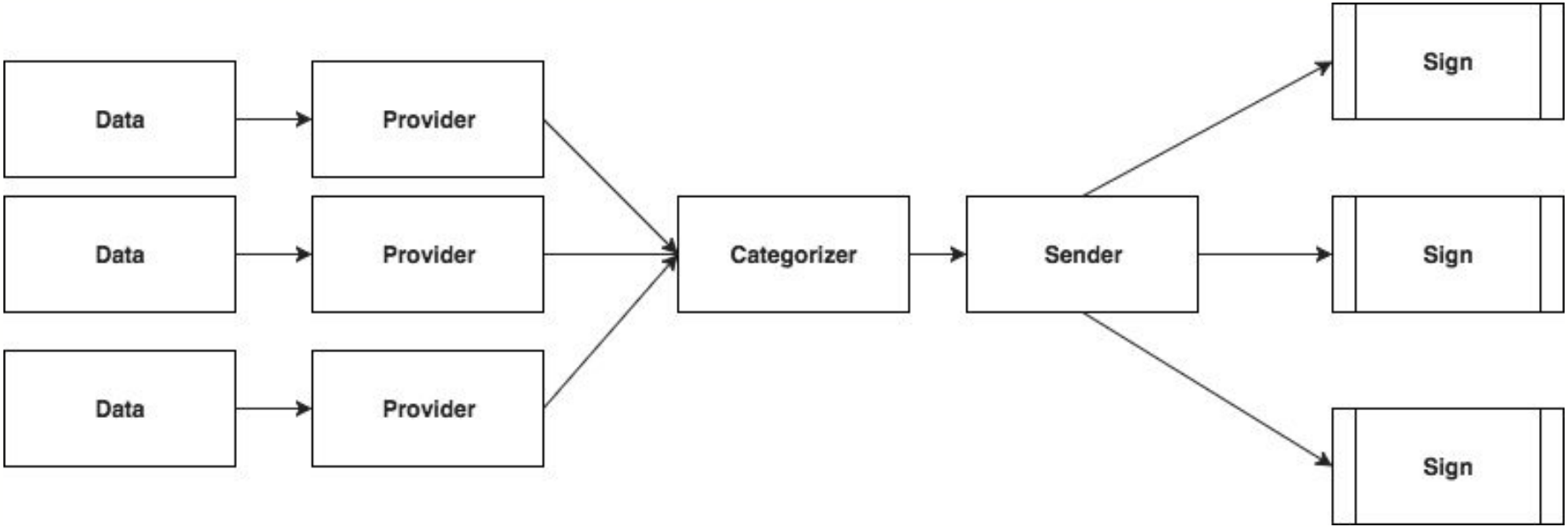
Entity Forms to Capture Relevant Info



Configuration Provider

- Sign Config
- Dynamic Updates to System

Pipeline

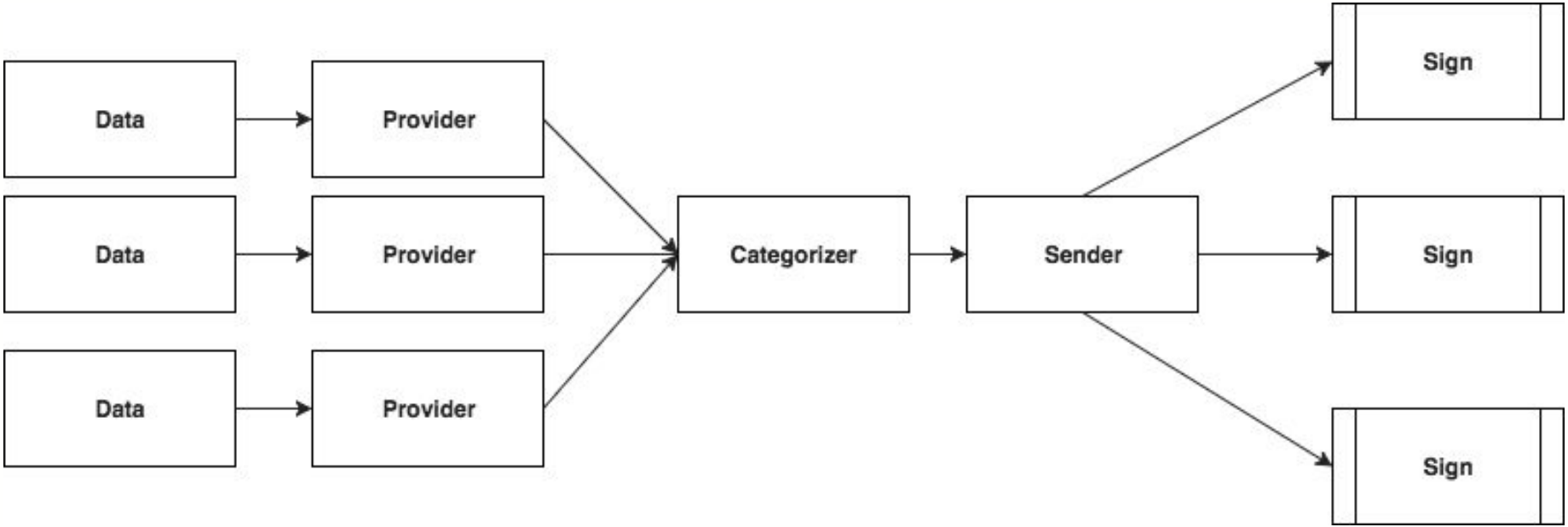


Categorizer

- Markers in Data + Drupal provided Entities
- Can be updated on the fly

Performance

Pipeline



Categorizer Requires Drupal

- Drupal has many moving parts
- Lots of structure / config leads to big / complex queries
- Big / Complex queries are SLOW

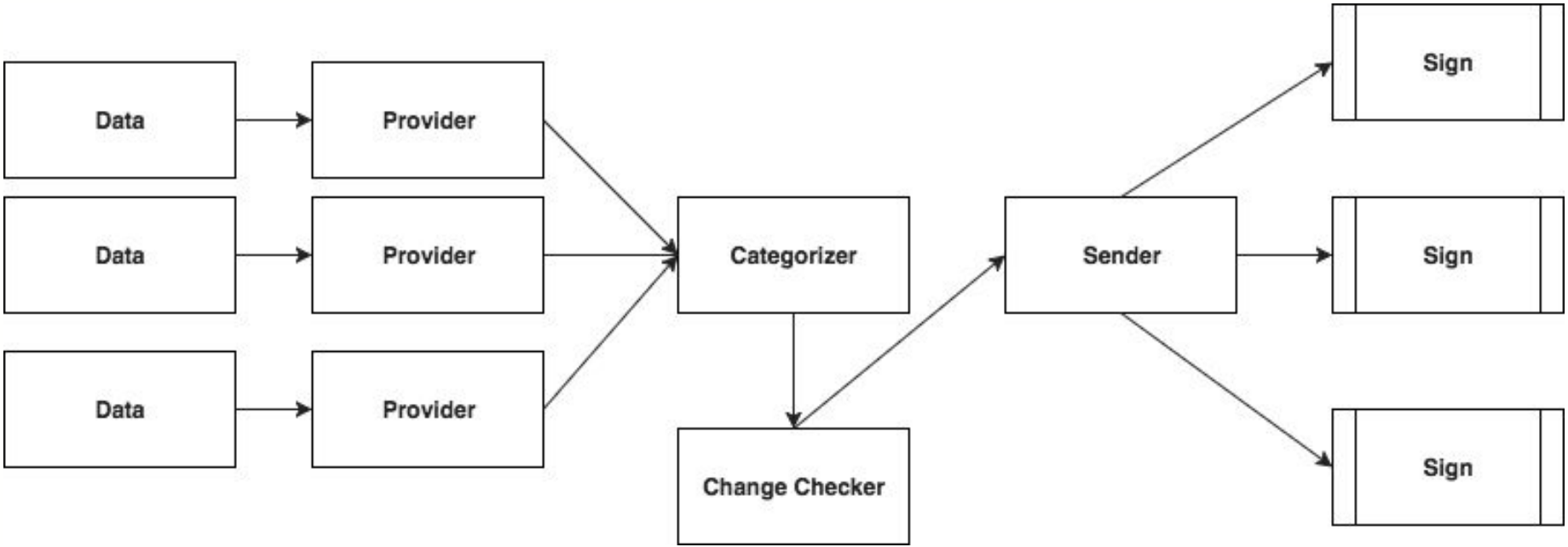
Cache EVERYTHING

- Pipeline accesses the cache
- Cache is regenerated only as needed
- Cache could be:
  - File Based
  - DB Based
  - Memcache Based

Did Anything Change???

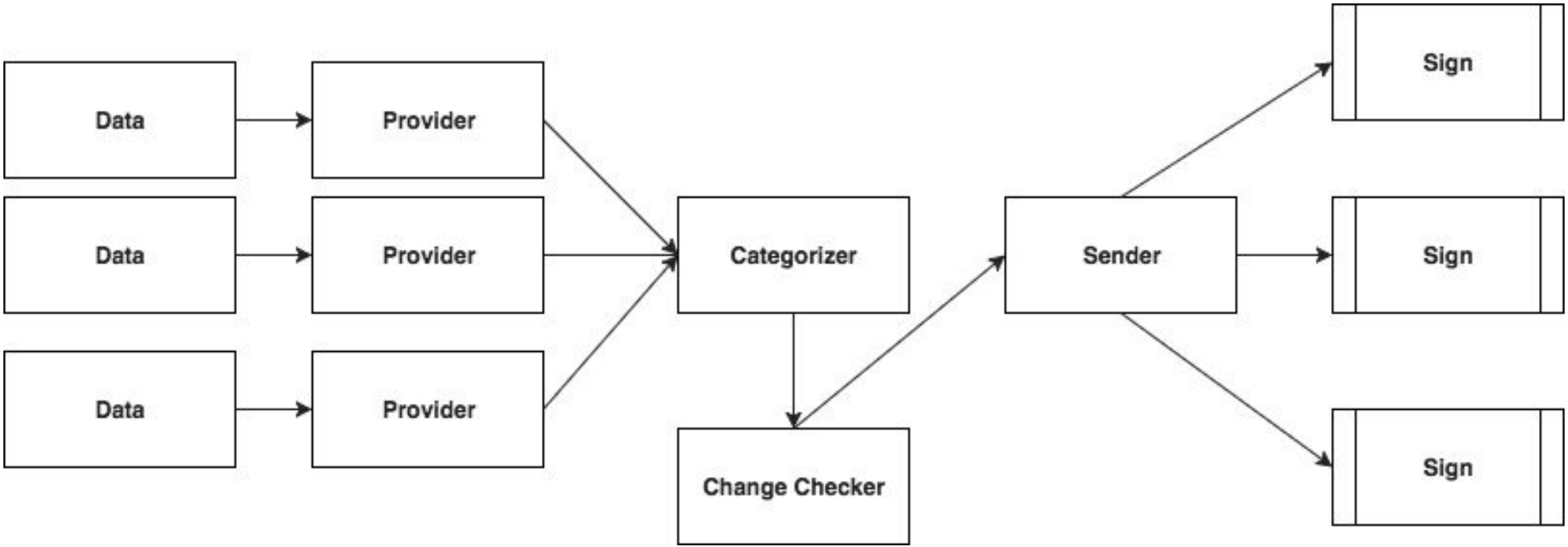
- Once Data is categorized… should we send it?

How Do You Show Data On A Sign?



Transmission

How Do You Show Data On A Sign?



Sending Data

- Normal Websites:
  - Requests occur as user's navigate
  - New data on page load
  - Asynchronous communication

Sending Data

- For Signs:
  - Persistent Connection
  - Fast
  - Lightweight

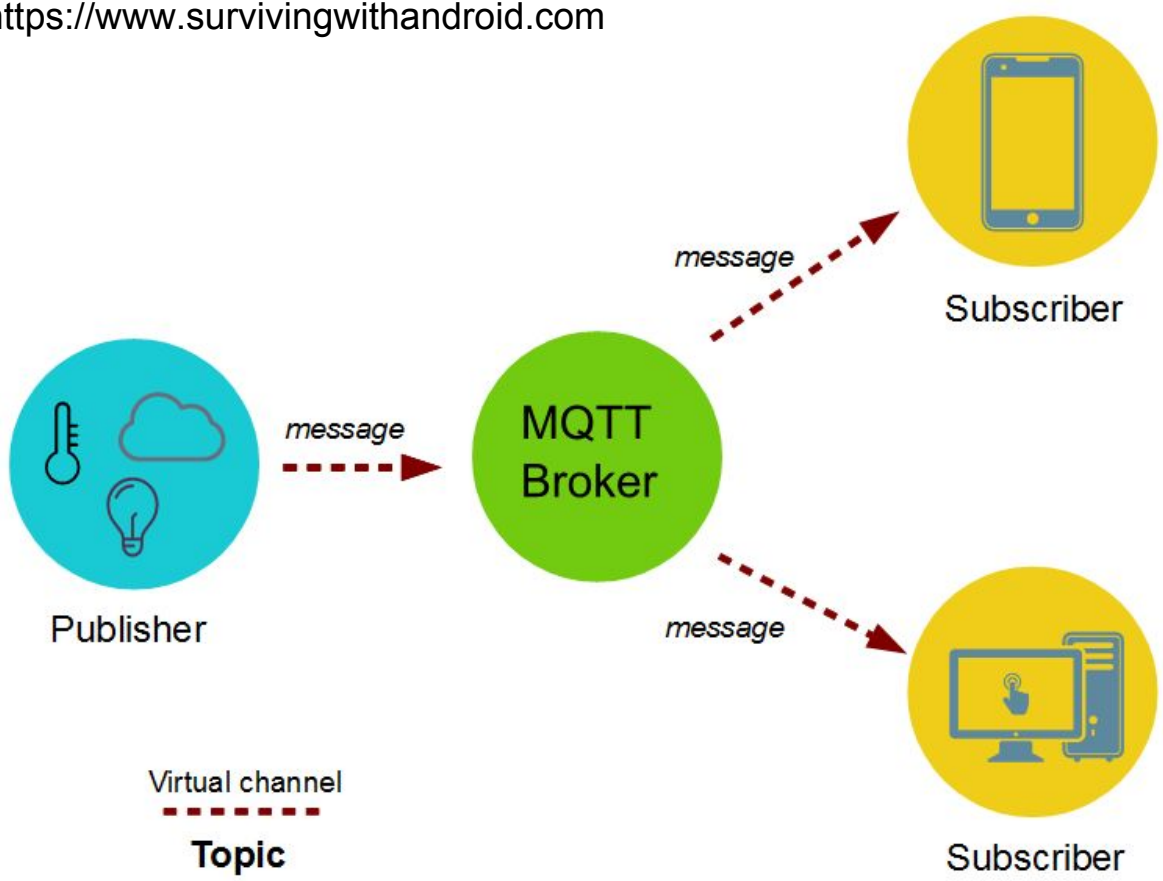
MQTT

Message Queuing Telemetry Transport

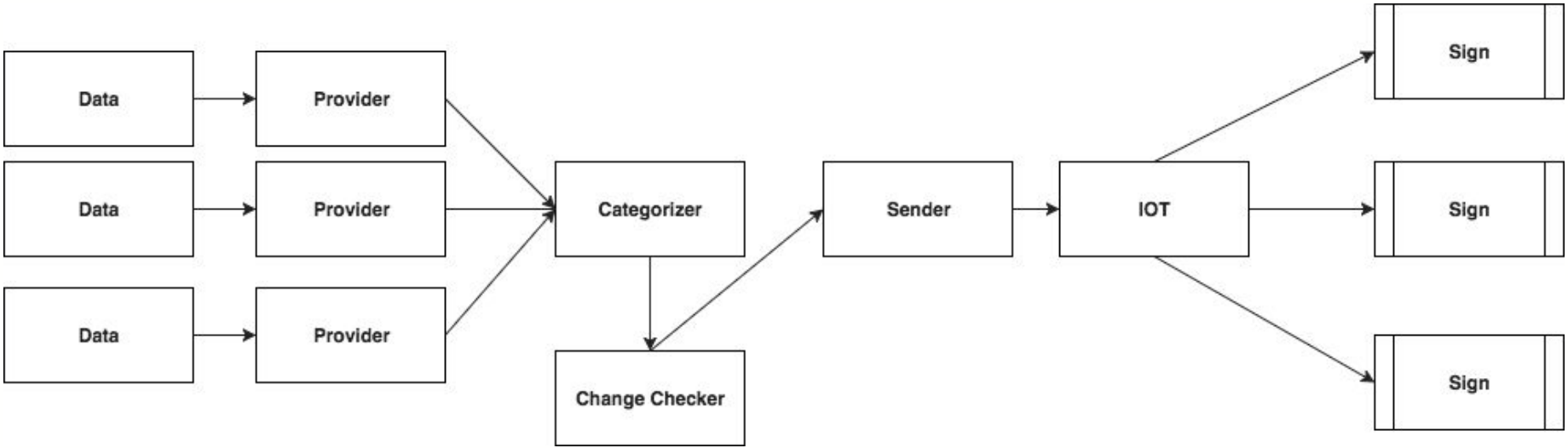
“publish-subscribe-based messaging protocol. It works on top of the TCP/IP protocol. It is designed for connections with remote locations where a "small code footprint" is required or the network bandwidth is limited. The publish-subscribe messaging pattern requires a message broker.”

https://en.wikipedia.org/wiki/MQTT

Image Source: https://www.survivingwithandroid.com



How Do You Show Data On A Sign?



Testing Plan (TDD)

Acquia Build and Launch Tool (BLT)

https://blt.readthedocs.io

Unit Testing

PHPUnit

- Ensures functionality at a method level
- Ensures classes / components function in a vacuum

(Not) Unit Testing

PHPUnit

- Bootstrap Drupal / Drush
- Execute Scripts / Caching

Functional Testing

Behat

- Ensure fields / entities are functional
- Test roles / permissions / workflows

Pipeline Testing

Behat

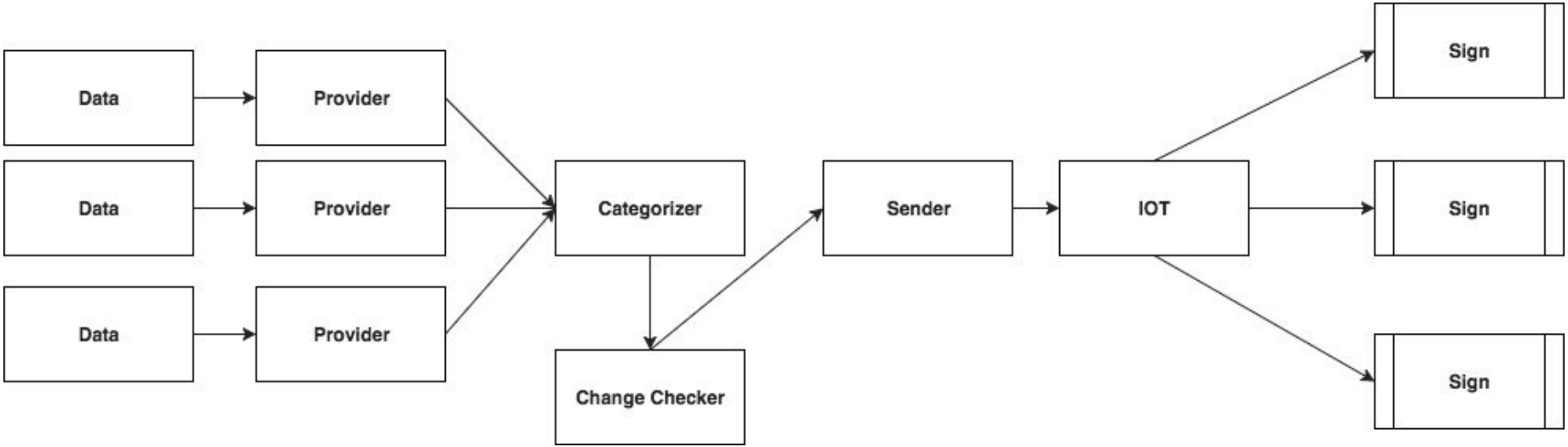
- Use Feature Context to execute elements of Pipeline
- Send data to IOT based on test script
- Retrieve data from IOT and compare with expected value(s)

SLOW but FIERCE

Automated tests take ~20 minutes…

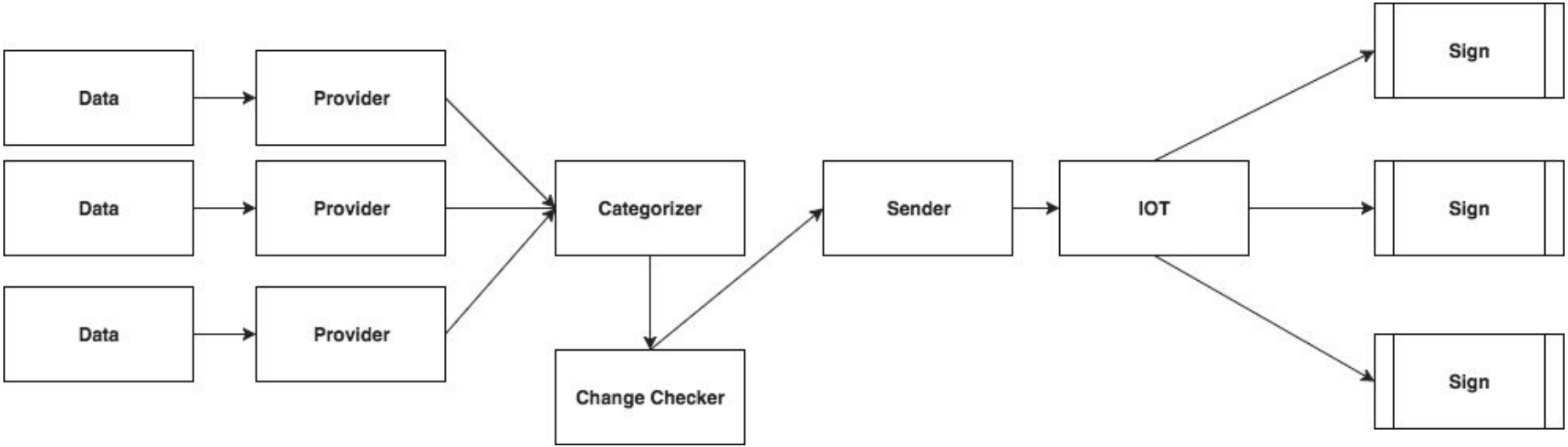
1. Communication with IOT
2. Results of Pipeline for ALL Data Providers
3. Queue Runners' Functionality
4. The usual unit + functional testing

Pipeline



Not much of a Drupal talk...

Pipeline



Drupal as a Data Provider

- Workflow
- Permissions / Roles
- WYSIWYG / Easy to Use Forms
- Contrib
- Security
- Media
- Views / Reporting

Drupal as a CMS

Questions