

GTFS: The Life of a Spec

Brian Ferris - Google



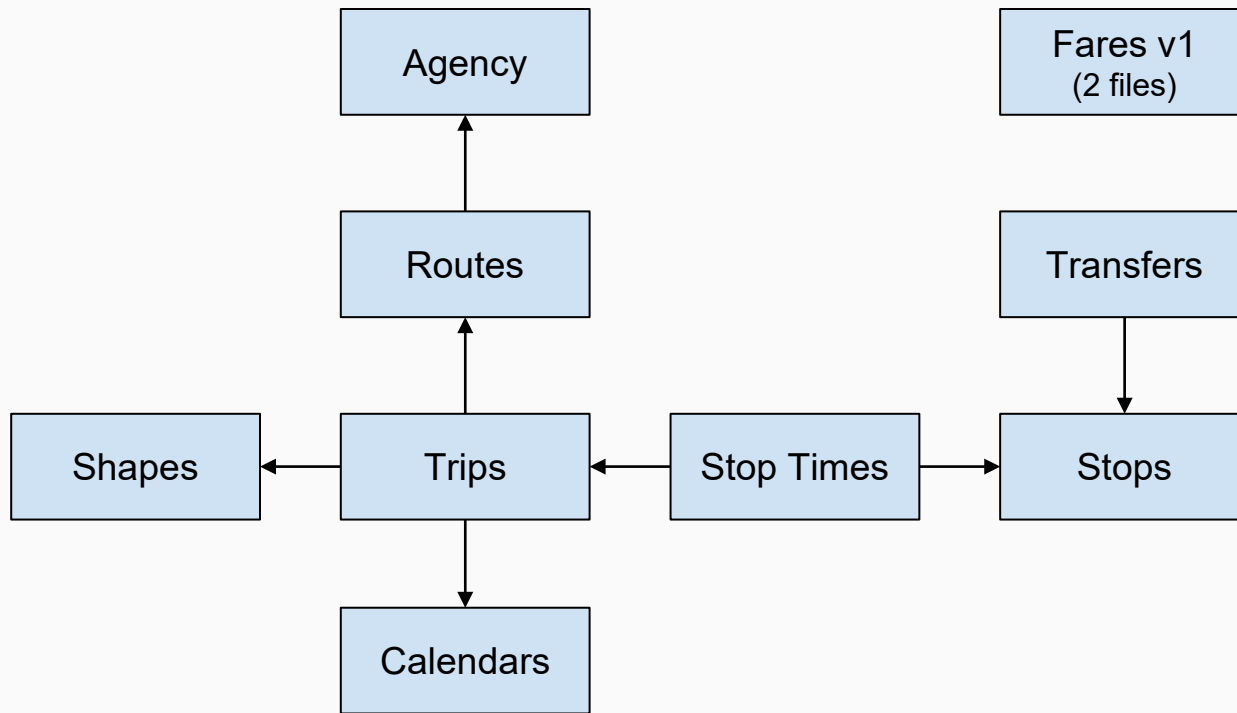
What is GTFS?

- The General Transit Feed Specification (GTFS) is:
 - a data specification
 - for exchanging public transit data
 - for use in rider information applications (like Google Maps)
- Google originally created the specification in 2006, in partnership with a number of US transit agencies.
- Today, GTFS is used by thousands of agencies, operators, applications, planners, researchers, enthusiasts (ahem transit nerds) around the world.
- <https://github.com/google/transit> + <https://gtfs.org/>

What does it look like?

- The spec specifies the names and fields of a series of CSV text files
 - e.g. `stops.txt`: `stop_id`, `stop_name`, `stop_lat`, `stop_lon`
- Each row provides data for a single entity (e.g. stops, routes, agencies)
- References between entities in different files
- All zipped up together and typically hosted somewhere on the web

The schema is... **complex**



But it started simpler.

How has GTFS changed over time?

- Over the years, we have:
 - Expanded the behavior of existing fields
 - Added new fields to existing files
 - Added new files
- Two aspects of change:
 - The **philosophy** of what to change
 - The **process** for how to change

The philosophy of GTFS

GTFS was born out of frustration with some earlier attempts at transit data standards, which were seen as:

- bloated, overly complex, difficult to evolve

Thus, few used them in practice.

The creators of GTFS wanted the evolution of the spec to be different.

GTFS Guiding Principles

<https://gtfs.org/schedule/process/#guiding-principles>

“The spec is about passenger information”

- GTFS is about passenger information
 - Not about operational information

Takeaway: Know what your spec **is** and, even more so, what it **isn't**.

- Helps keep things focused over time

“Feeds should be easy to create and edit”

- Theoretically, you can edit GTFS with a spreadsheet
 - (some agencies still do)
- There is a lot of tooling and vendor support these days
 - But was not always the case
- There are a lot of transit agencies in the world
 - Most do not have in-house software engineering teams
 - Comparatively fewer groups consuming GTFS; often more technically sophisticated

Takeaway: Make things easier for the stakeholders who need it

“Feeds should be easy to parse”

- While consumers of GTFS may be more technically sophisticated, we don't want to make their lives too difficult
- Some data representations are easier to work with than others

Takeaway: Think about what data consumers can practically use when designing a data format

“Changes to the spec should be backwards-compatible”

- A feed published under a previous version of GTFS is still valid GTFS
 - Reflects that data producers (transit agencies) are limited in their ability to update software and processes
- But new features can still be added to the spec
 - Puts burden on data consumers to update software and processes

Takeaway: Changes come with upgrade costs so be deliberate

- Think about who pays those costs

“Speculative features are discouraged”

- GTFS encourages that any new feature have both a producer and a consumer willing to use it
- Avoids the dual problems of specs growing unnecessarily:
 - Data that no one actually wants to use
 - Data that no one can actually provide

Takeaway: It's ok to think about the future evolution of a spec, but try to adopt features that will actually get used.

GTFS Change Process

<https://gtfs.org/schedule/process/>

GTFS Change Process

- Changes are adopted via a consensus-based voting process
- Anyone can participate, both in proposing changes and voting on them
- Consensus means that a single no-vote is enough to veto a proposal
 - “No” votes need to include constructive feedback

It is a messy, organic, sometimes-chaotic process that somehow sort of works.

Facilitating the process

- Large changes are hard; they require energy and coordination
- In early days, individual volunteers (like me) oversaw the change process
- These days, we have help from MobilityData
 - Non-profit organization focused mobility data standards
 - Funded by members of the transit data community
- MobilityData serves as the main facilitator for the GTFS change process
 - Facilitates online discussions, confirming voting results, updating spec
 - Coordinates working groups for large discussions (e.g. GTFS-Flex, Fares v2)
- Coordination through

Questioning the process

- How do you encourage a diversity of voices in the evolution of the spec?
 - Involvement takes awareness and time from stakeholders
- How do you encourage an efficient process?
 - A diversity of voices makes it harder to reach consensus
- Is consensus the right model?
 - Can one bad actor derail the process? Majority voting? Who gets to vote?
- Iteration during spec development vs implementation costs?
 - Changing a proposed schema requires early-adopters to update
- What about extensions?
- Would we consider a major revision of GTFS?

Future of the process

- MobilityData leading a Working Group on the Change Process itself
 - <https://github.com/google/transit/issues/436>
 - (I'm a participant)
- GTFS is 18 years old at this point
 - Still capable of big changes (GTFS-Flex, Fares v2)
 - But becoming a mature adult as well
 - Change is more deliberate because the stakes are higher

Questions?