# Towards a next generation time series format standard

and some IRIS specifics...

# It was 1988… SEED 2.0 landed

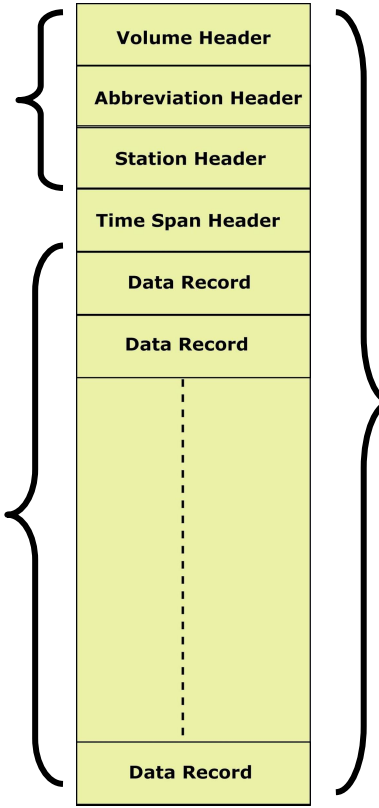
- Governed by Working Group 2 of the International Federation of Digital Seismograph Networks (FDSN)
- Format for time series and comprehensive related metadata
- Primary purpose is exchange and archiving
- Extensible, kinda, sorta
- Very successful!

**dataless SEED**

**(ASCII)**

**mini-SEED**

**(binary + ASCII)**




**(full) SEED volume**

StationXML

???

# Limitations of miniSEED 2.4

- Fixed width identifiers (aka codes), with a single naming scheme
- Many awkward-isms (e.g. record length) or retroactive confirmation (byte order)
- Bizarrely complex time presentation:
    - Base + 2 x microseconds + correction + corr flag (+ leap second)
- Custom, blockette structure
- Fixed record lengths, blessing and curse

# Identifiers need expansion

- Not ready for Large-N
- Temporary network codes are a kludge, painful when reused
- Out of instrument identifier codes
- No sensible way to ID derived or synthetic data explicitly

Sweetwater: 2,268 sensors in a 7x km$^2$ area

Sagebrush Flats: 1,108

# FDSN steps towards a next generation time series format

**Mid-2016**: IRIS spearheads a process to develop a next generation miniSEED, starting with a “strawman” as a nucleus for discussion

- Significant feedback received from a broad range of stakeholders

**Early-2017**: Focussed Working Group II meeting in Utrecht, the Netherlands

**Mid-2017**: FDSN General Assembly, report from focus meeting

**Early-2018**: Working Group II process re-initiated to collaboratively determine requirements for a next generation time series specification

# Next Generation Format (NGF) requirements

- Use a time series identifier that is a single, mandatory but flexible name-spaced string for the in the approximate form:

  `FDSN:<network>_<station>_<location>_<channel>`

- Parsable without resolving with heuristic checks. Information not in an initial, fixed header should be encoded in a standardized format
- Data loss is acceptable, a few cases of very rarely or never used structures
- Include a CRC for the whole record
- Include a format version
- Include a data publication version
- Allow arbitrary extra headers in a key-value data structure
- Allow variable record lengths
- Nanosecond time resolution
- Double precision sampling rate
- Historical problem areas fixed, e.g. set fixed byte order, no unapplied time corr., etc.

# Next steps for FDSN Working Group 2

- Review proposals when they are submitted
  - Timeline and details TBD

- Initiate discussion on time series identifiers
  - Broad participation desired

*It is not too late to participate and contribute!*

# Expansion of time series identifiers

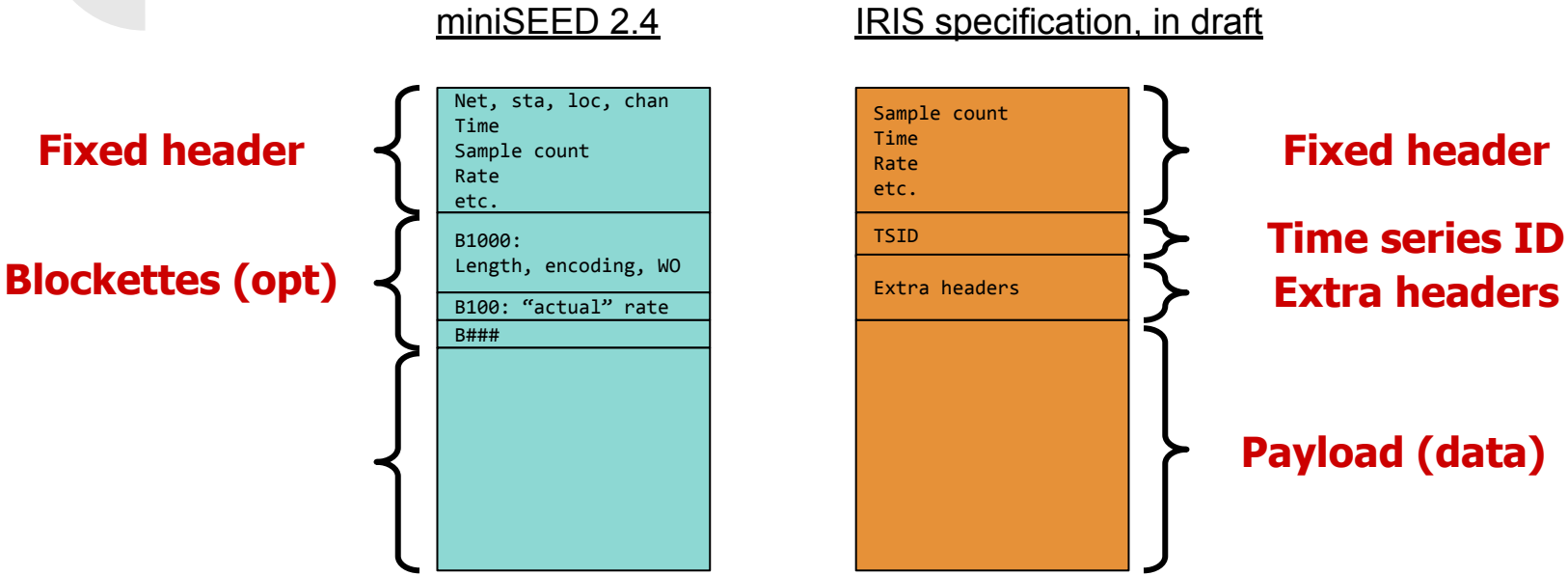
## Likely the most impactful change

| Currently | Potential expansion? |
| --- | --- |
| Network code (2 characters) | 8 characters |
| Station code (5 characters) | 8 characters |
| Location code (2 characters) | 8 characters |
| Channel code (3 individual codes) | 3 or 4 or more characters |
| | New codes? |

Form of the new identifier?

```
FDSN:<network>_<station>_<location>_<channel>
```

# The IRIS specification




Plans: spend the next N months doing technical evaluation, developing resources and then present the specification to the FDSN for consideration.