# Principles of **Data Usability**

1. Share your data.

# Principles of **Data Usability**

1. Share your data.
2. Provide metadata.

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.
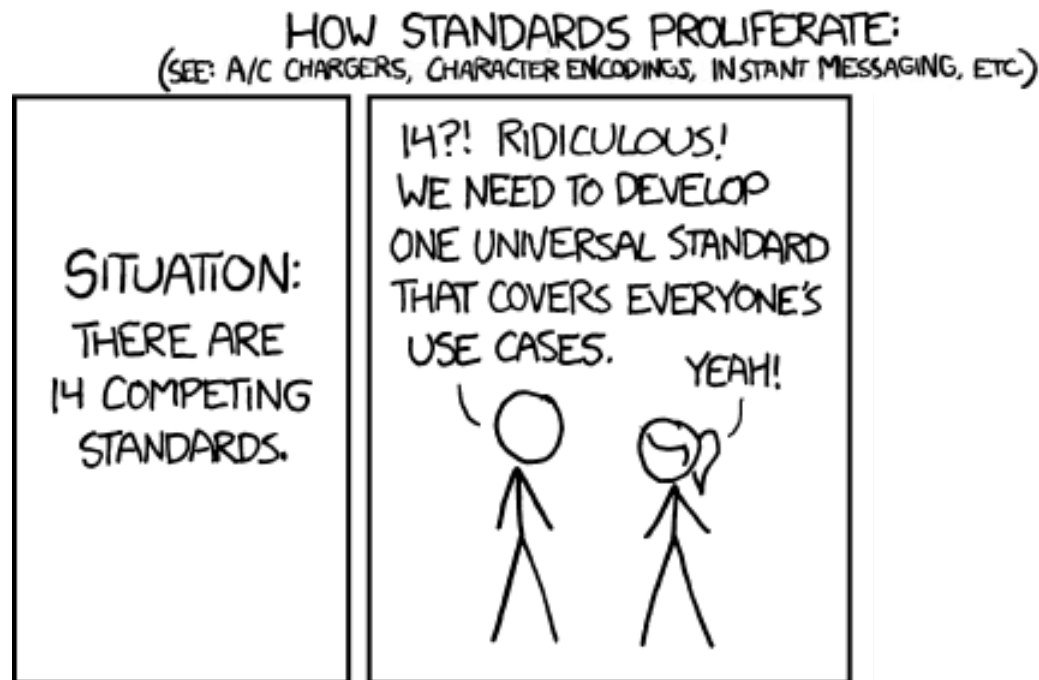
# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.



HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

Randall Munroe, *xkcd*.

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.



Randall Munroe, *xkcd*.
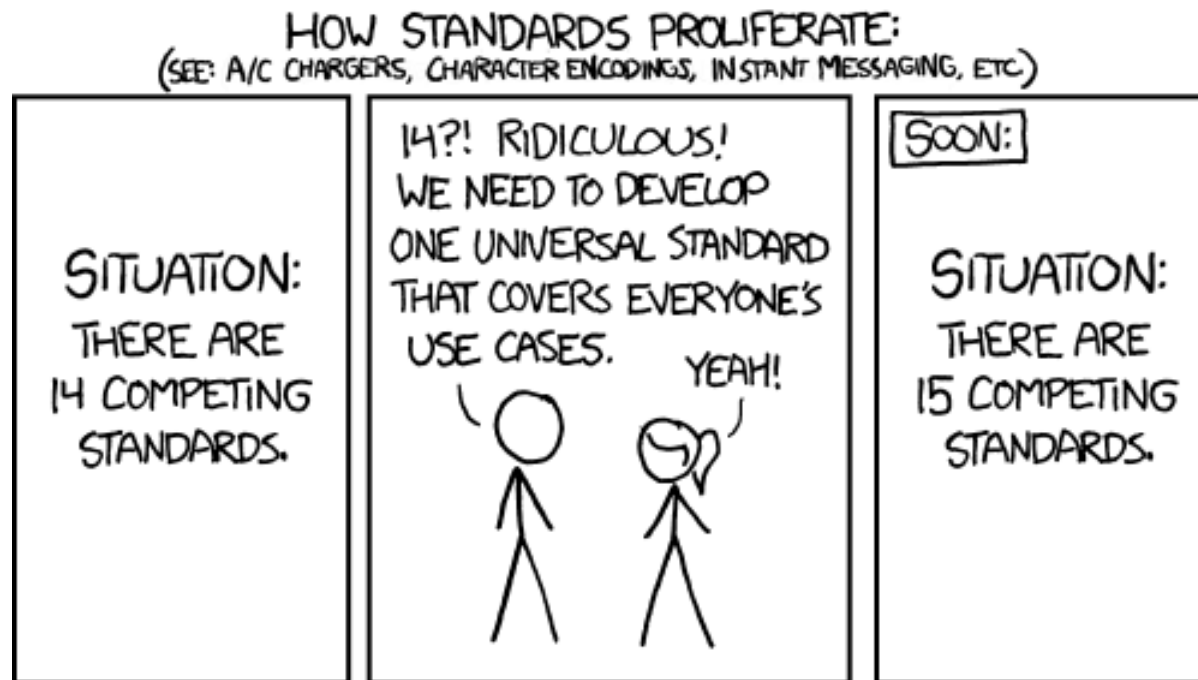
# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.



Randall Munroe, *xkcd*.

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.

| Format Name | Description |
| --- | --- |
| RAW | Sequence format that doesn't contain any header. |
| FASTA | Default sequence format.  Contains header line and sequence: |
| | `>name` |
| | `AGCTGTGTGGGTTGGTGGGTT` |
| PIR | Sequence format that's similar to FASTA but less common |
| MSF | Multiple sequence alignment format |
| CLUSTAL | Multiple sequence alignment format (works with T-Coffee) |
| TXT | Text format |
| GIF, JPEG, PNG | Graphic formats |
| PDF, DOC | Binary format |

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.

5. Use good null values.

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.

5. Use good null values.

6. Make it easy to combine your data with other datasets.

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.

5. Use good null values.

6. Make it easy to combine your data with other datasets.

7. Perform basic quality control.

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.

5. Use good null values.

6. Make it easy to combine your data with other datasets.

7. Perform basic quality control.

8. Use an established repository.

# Principles of **Data Usability**

1. Share your data.

2. Provide metadata.

3. Provide an unprocessed form of the data.

4. Use standard data formats.

5. Use good null values.

6. Make it easy to combine your data with other datasets.

7. Perform basic quality control.

8. Use an established repository.

9. Use an established and open license.
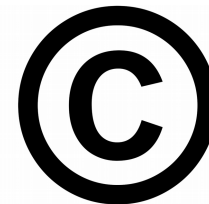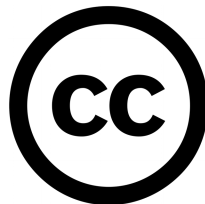
# Bioinformatics **Common Formats**

| Format | Description |
|---|---|
| RAW | Sequence format without header. |
| FASTA | Sequence format with header line and sequence: |
| | `>name AGCTGTGTGGGTTGGTGGGTT` |
| PIR | Sequence format similar to FASTA |
| MSF | Multiple sequence alignment format |
| CLUSTAL | Multiple sequence alignment format |
| TXT | Text format |
| GIF, JPEG, PNG | Graphic formats |
| PDF, DOC | Binary format |