

Podcasting Data Formats (Part 1)

All Hail to the Detail

Tim Pritlove

<http://tim.geekheim.de/>
PodCamp, 14. Januar 2007

Publication Formats

- Audio Recordings
- Video Recordings
- Photos
- Text Documents
- Syndication (RSS)

Codecs and Containers



- Often confused or confusingly used
- Codecs encode audible or visual information
- Containers organize encoded information in tracks
- Most file formats are container formats

Containers and Tracks

CODEC

Audio

CODEC

Video

CODEC

Text

Container

Popular Container Formats

- MPEG-1 (.mpg, .mp3)
- MPEG-4 (.mp4, .m4v, .m4a, .3gp)
- OGG (.ogg)
- QuickTime (.mov)
- Windows Media (.wma, .wmv)
- Audio Video Interleave (.avi)

The AVI Disaster

- Underspecified Container Format
- Abandoned by its creator
- Widespread use but many ways of using it
- Many corrupt or inconsistent files
- Flash Video to our help?

MPEG-4

The problematic saviour

- Rather well-defined standard
- Broad support by Apple
- Part of many future technologies
- Heavily patented
- License fees
- DRM options

Popular Video Codecs

- MPEG-1 Video
- MPEG-4 Video, XVID
- MPEG-4 Advanced Video Codec (H.264)
- Windows Media Video Codecs (WMV)
- Flash Video (Sorenson H.263)
- Theora

The DIVX/XVID Confusion

- Alternative Codec based on preliminary Microsoft MPEG-4 development
- MPEG-4 incompatible
- Commercial development
- XVID is an open source MPEG-4-compatible version alternative
- XVID is the pirate's friend

Which format to choose?

- MPEG-4 is the de-facto Podcasting standard
- Flash Video is the de-facto web standard
- WMV even worse than MPEG-4 in terms of licensing, patents and DRM
- Theora is totally free, pretty good but widely unused and has a lack of tools
- Battle not yet decided

Popular Audio Codecs

- MPEG-1 Audio Layer III (MP3)
- MPEG-4 Advanced Audio Codec (AAC)
- Windows Media Audio (WMA)
- Vorbis
- Free Lossless Audio Codec (FLAC)

Audio Encoding Parameters

- Sampling Rate (22/44.1/48 kHz)
- Bitrate (16–256 kBit/s)
 - Constant Bitrate Encoding
 - Variable Bitrate Encoding
- Mono and stereo
- Quality should come first

Podcasting Data Formats (Part 2)

All Hail to the Detail

Tim Pritlove

<http://tim.geekheim.de/>
PodCamp, 13. Januar 2007

ID3 Meta Data

- In-Band Meta Data
- Title, Artist, Year, Comments
- Composer
- Lyrics
- Artwork

Syndication

- RSS 2.0 is the Podcasting standard
- Atom 1.0 is not yet used
- XML documents
- iTunes and Yahoo extensions

Syndication

- RSS 2.0 is the Podcasting standard
 - Loosely defined XML documents
- Atom 1.0 is much better
 - Strictly defined XML documents
 - No wide support in sight
- iTunes and Yahoo extensions

RSS 2.0

Channel Description

- `<title>`
- `<itunes:subtitle>`
- `<description>`
- `<link>`
- `<itunes:summary>`
- `<Language>`

RSS 2.0

Channel Ownership

- `<generator>`
- `<itunes:owner>`
 - `<itunes:author>`
 - `<itunes:email>`

RSS 2.0

Channel Additions

- `<category>`
- `<itunes:category>`
- `<itunes:keywords>`
- `<itunes:image>`
- `<itunes:explicit>`
- `<itunes:block>`

RSS 2.0

Document Structure

```
<rss version="2.0"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:content="http://purl.org/rss/1.0/modules/content/"
  xmlns:itunes="http://www.itunes.com/dtds/podcast-1.0.dtd"
  xmlns:xhtml="http://www.w3.org/1999/xhtml" >

  <channel>
    ...
  </channel>

  <item>
    ...
  </item>
  ...
</rss>
```

RSS 2.0

Channel Description

```
<channel>
  <title>Chaosradio</title>
  <itunes:subtitle>Das monatliche Live-Talk-Radio auf Fritz</itunes:subtitle>
  <link>http://chaosradio.ccc.de/chaosradio.html</link>
  <description>Chaosradio ist der unterhaltsame Live-Talk-Radio-Klassiker des Chaos Computer Clubs aus Berlin und eines der ältesten Tech-Radios überhaupt. Chaosradio informiert seit 1995 über wechselnde Themen rund um Technologie und Gesellschaft.</description>
  <itunes:summary>Chaosradio ist der unterhaltsame Live-Talk-Radio-Klassiker des Chaos Computer Clubs aus Berlin und eines der ältesten Tech-Radios überhaupt. Chaosradio informiert seit 1995 über wechselnde Themen rund um Technologie und Gesellschaft.</itunes:summary>
  <language>de</language>
```

RSS 2.0

Channel Description

```
<generator>Chaosradio Web Site and Feed Generator XSLT
Script</generator>
<itunes:author>Chaos Computer Club</itunes:author>
<itunes:owner>
  <itunes:name>Chaos Computer Club</itunes:name>
  <itunes:email>chaosradio@ccc.de</itunes:email>
</itunes:owner>
<itunes:image href="http://chaosradio.ccc.de/chaosradio-
logo-300x300.jpg"/>
<category>Technology</category>
<itunes:category text="Technology"/>
<itunes:keywords>ccc,hacking,computer,technology,society,
chaosradio,chaos,discord,confusion</itunes:keywords>
<itunes:explicit>no</itunes:explicit>
</channel>
```

RSS 2.0

Item Definitions

- `<title>`
- `<author>`, `<itunes:author>`
- `<description>`,
`<itunes:subtitle>`,
`<itunes:summary>`
- `<pubDate>`, `<dc:date>`

RSS 2.0

Item Content

- `<description>`
- `<content:encoded>`
- `<xhtml:body>`
- `<enclosure>`
- `<itunes:duration>`

RSS 2.0

Permalinks

- `<link>`
- `<guid isPermaLink="false">`
 - Use serial numbers to update postings
 - Separate number from URL via '#'

RSS 2.0

Item Description

```
<item>
  <title>CR119 Instant Messaging</title>
  <author>Chaos Computer Club</author>
  <itunes:author>Chaos Computer Club</itunes:author>
  <itunes:subtitle>Realtime Kommunikation im Internet</itunes:subtitle>
  <description>Realtime Kommunikation im Internet</description>
  <enclosure type="audio/mpeg" length="168513607"
    url="http://chaosradio.ccc.de/archive/chaosradio_119.mp3" />
  <itunes:duration>02:55:27</itunes:duration>
  <guid isPermaLink="false">http://chaosradio.ccc.de/cr119.html#0001</guid>
  <link>http://chaosradio.ccc.de/cr119.html</link>
  <category>instant messaging</category>
  <category>xmpp</category>
  <category>jabber</category>

  <pubDate>01 Dec 2006 13:30:00 +0100</pubDate>
  <dc:date>2006-12-01T13:30+01:00</dc:date>

  <itunes:explicit">no</itunes:explicit>
```

RSS 2.0

Item Description

<itunes:summary>Chaosradio CR119
Sendung vom: 30.11.2006, 20:00 Uhr
<http://chaosradio.ccc.de/cr119.html>

Instant Messaging

Realtime Kommunikation im Internet

Team: Tim Pritlove, Florian Holzhauer, Holger Klein

Instant Messaging (IM) ist nichts neues, doch setzt es in dieser Zeit zum großen Sprung an. Zahlreiche Diensteanbieter wie Google, GMX oder Web.de liefern heute zur kostenlosen E-Mail-Inbox auch Instant Messaging-Dienste dazu und treiben ein einheitliches Messaging-Netzwerk voran. IM wird zur SMS des Internets - ist aber deutlich billiger und vor allem deutlich flexibler.

</itunes:summary>

RSS 2.0

Item Description

```
<xhtml:body>
```

```
<p>
```

```
Instant Messaging (IM) ist nichts neues, doch setzt es in  
dieser Zeit zum großen Sprung an. Zahlreiche Diensteanbieter  
wie Google, GMX oder Web.de liefern heute zur kostenlosen E-  
Mail-Inbox auch Instant Messaging-Dienste dazu und treiben ein  
einheitliches Messaging-Netzwerk voran. IM wird zur SMS des  
Internets - ist aber deutlich billiger und vor allem deutlich  
flexibler.
```

```
</p>
```

```
<p>Team: Tim Pritlove, Florian Holzhauer, Holger Klein</p>
```

```
</body>
```

Podcast Feed Redirection

- HTTP Status Code “301 Moved Permanently” advertises URL change
- Supported by iTunes and other RSS readers
- Supported by iTunes Music Store
- Bad support by many other portals
- Feed ownership via Podcast feed