

# The Mars Project — PDF in XML

Matthew Hardy

Computer Scientist

Adobe Systems Incorporated



# Outline

- Brief introduction to PDF
  - History
  - Features
  - Representation
- Mars
  - What is it?
  - Why?
  - How?
  - Layout
- Demo
- Summary
- Questions

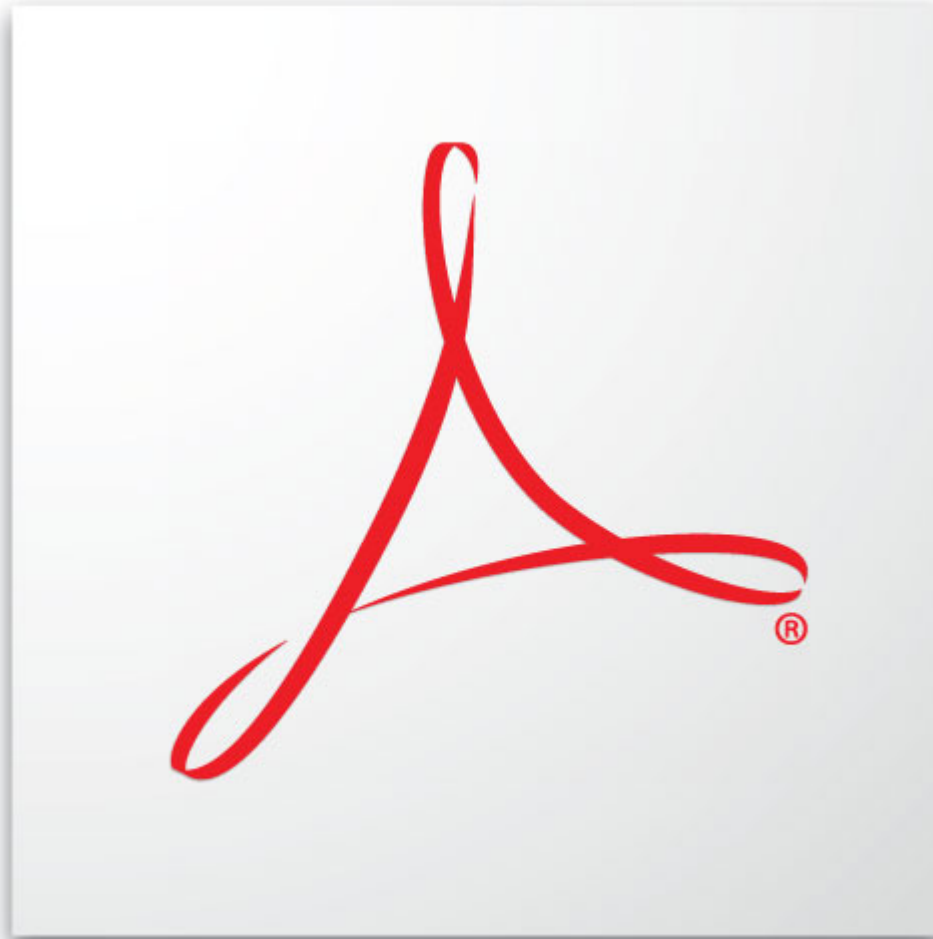
# Brief history of PDF

- Created in 1990 and released in 1993
- Device/Platform neutral Page Description Language
- High fidelity precision text layout
- High end graphic features
- Color management
- ISO standards PDF/X, PDF/A, PDF/E and soon PDF
- Extended to meet needs beyond its initial conception

# PDF Capabilities

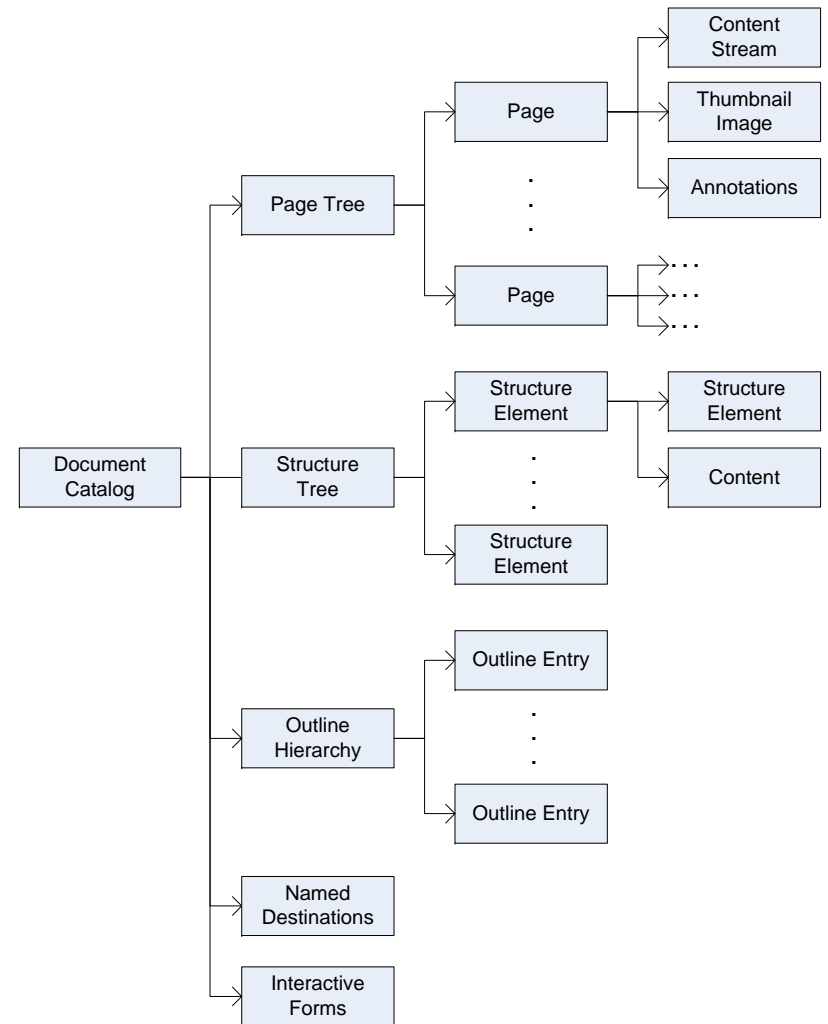
- Document Navigation (e.g. bookmarks, hyperlinks, etc.)
- Fillable Forms and XFA
- Logical Structure
- Accessibility/Content Access
- Video and Sound
- Annotations/Collaboration/Review
- Layers and 3D objects
- Signatures/Security
- Document Packages

# Demo of PDF — Adobe Acrobat®



# File Structure of PDF

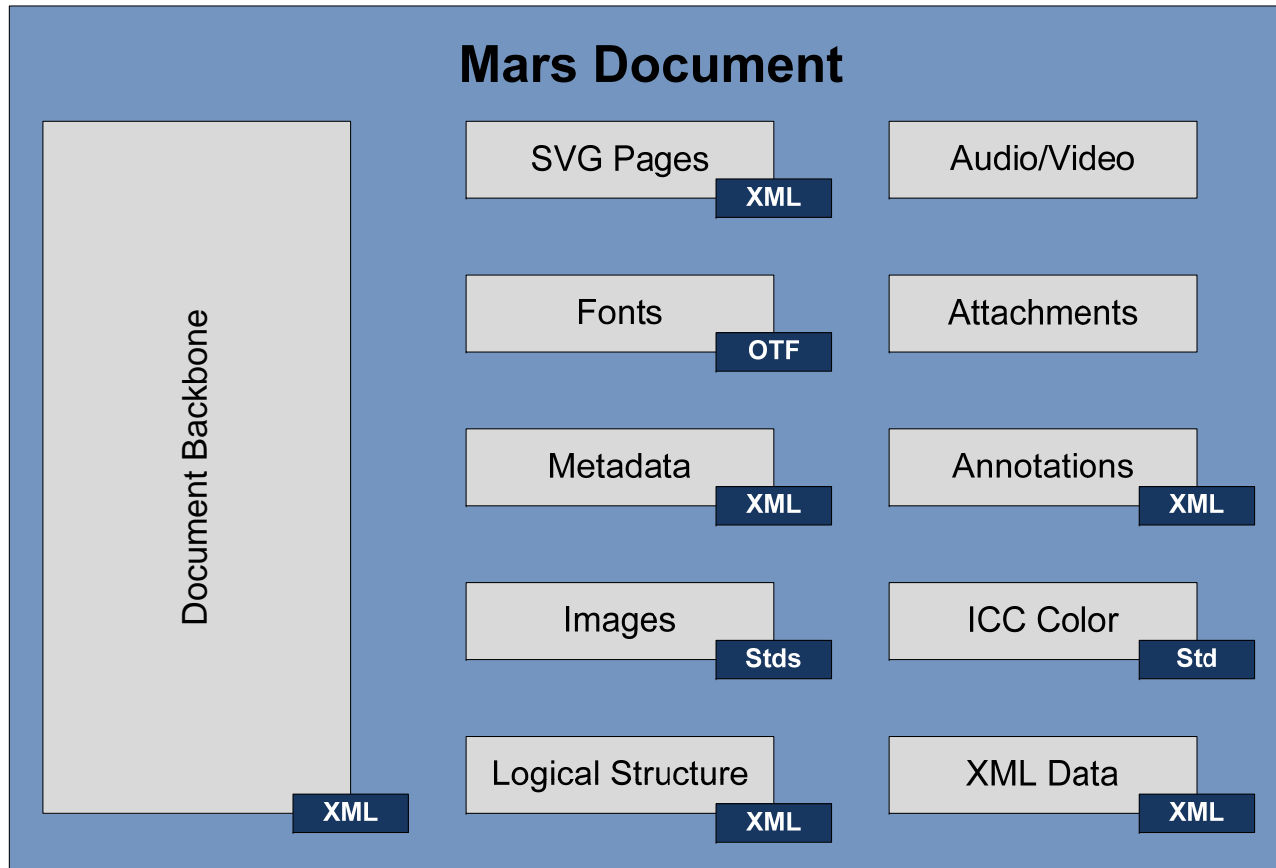
- PDF is based on an object model
  - Non-scalar
    - Dictionaries
    - Arrays
    - Streams
  - Scalar
    - Strings
    - Integers
    - Floats
    - Boolean
    - Names
- Objects combined to form a tree or graph
- Sub-trees in the PDF represent different aspects of the document



# What is Mars?

- Mars is an alternative representation for PDF documents
  - Packaged format using Zip/UCF to store components
  - XML is used to represent the various structural components of PDF
  - Industry standard formats are used for binary data (e.g. JPEG, PNG, ICC, OpenType, etc.)
- Feature parity with PDF
  - Capable of representing the same content as PDF
  - Looks identical to the end user
  - Complete representation of PDF

# More Mars

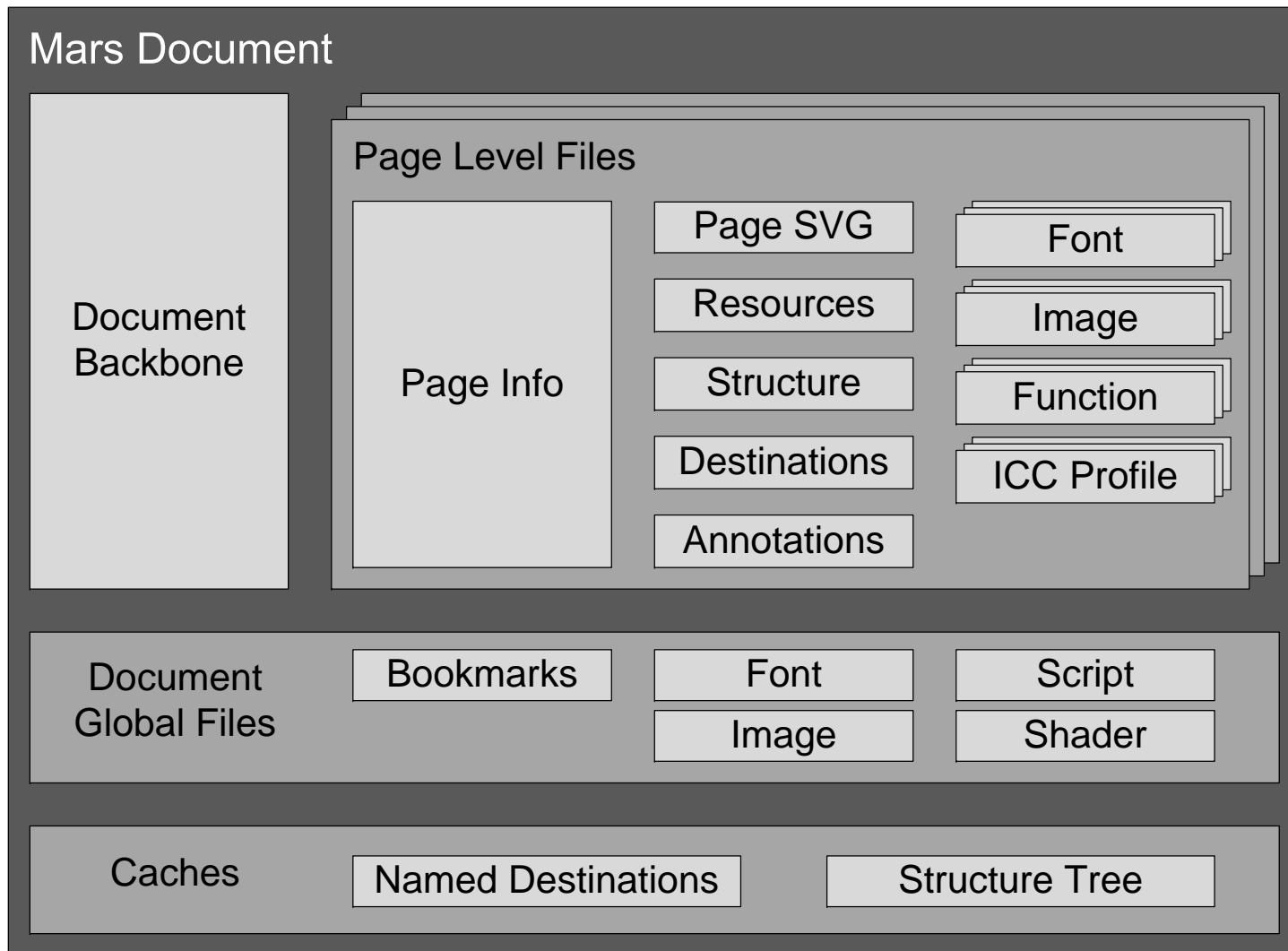




# Why are we doing Mars?

- If we were creating PDF today, we would implement it like Mars
  - Leverage current technology and IT infrastructure and know-how
  - Takes advantage of existing standards
- Developer friendly
  - Easy to integrate into modern workflows
  - Use of many existing tools
- Address requirements of government and business
  - Governments like/mandate use of open standards
  - Businesses follow and also like to standardize IT technology
- So to be well positioned for the next 15 years...
  - We need to move our technology base forward

# Mars Layout



# An Example

```
%PDF-1.4
1 0 obj
<<
/Type /Catalog
/Pages 2 0 R
>>
endobj
```

```
2 0 obj
<<
/Type /Pages
/Kids [3 0 R]
/Count 1
>>
endobj
```

```
3 0 obj
<<
/Type /Page
/Parent 2 0 R
/Resources 4 0 R
/Contents 5 0 R
/MediaBox [0 0 612 792]
>>
endobj

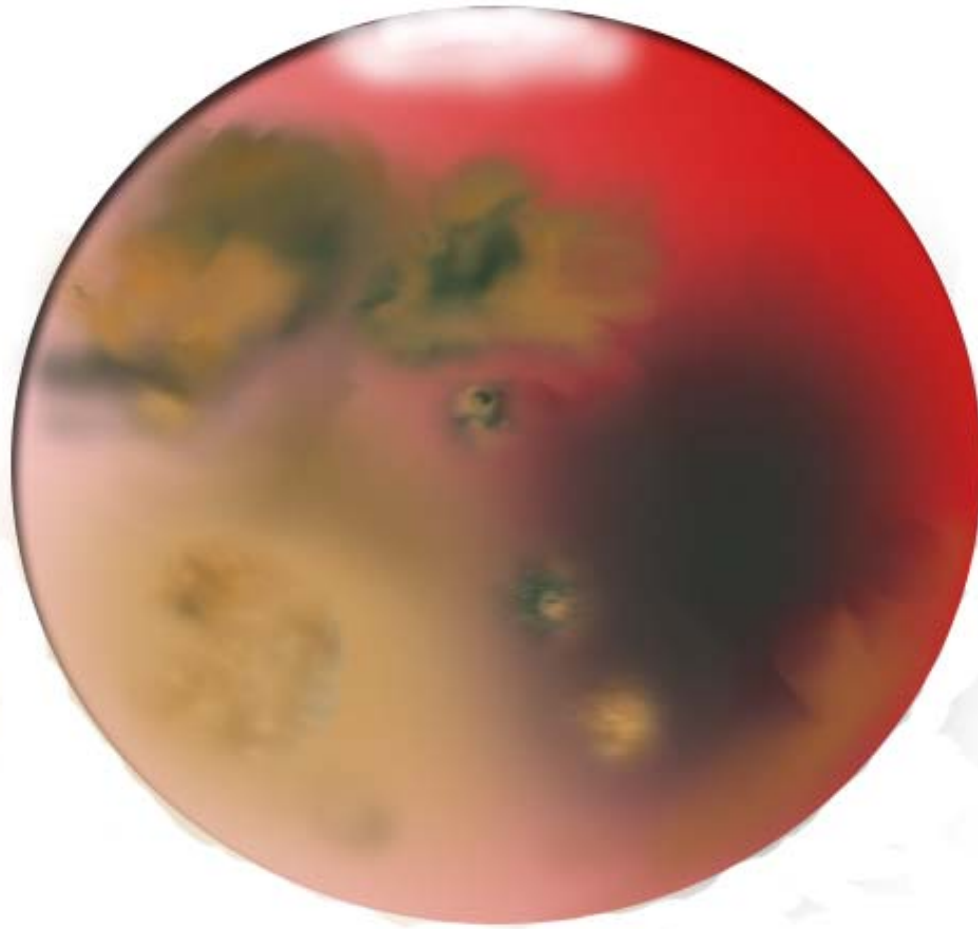
4 0 obj
<<
/ProcSet [/PDF /Text]
/Font <</F1 6 0 R >>
>>
endobj
```

```
5 0 obj
<<
/Length 115
>>
stream
BT
/F1 24 Tf
1 0 0 1 260 600 Tm
(Hello World) Tj
1 0 0 1 260 550 Tm
(Goodbye Universe!) Tj
ET
endstream
Endobj
```

```
6 0 obj
<<
/Type /Font
/Subtype /Type1
/Name /F1
/BaseFont /Helvetica
>>
endobj

trailer
<<
/Root 1 0 R /Size 7
>>
```

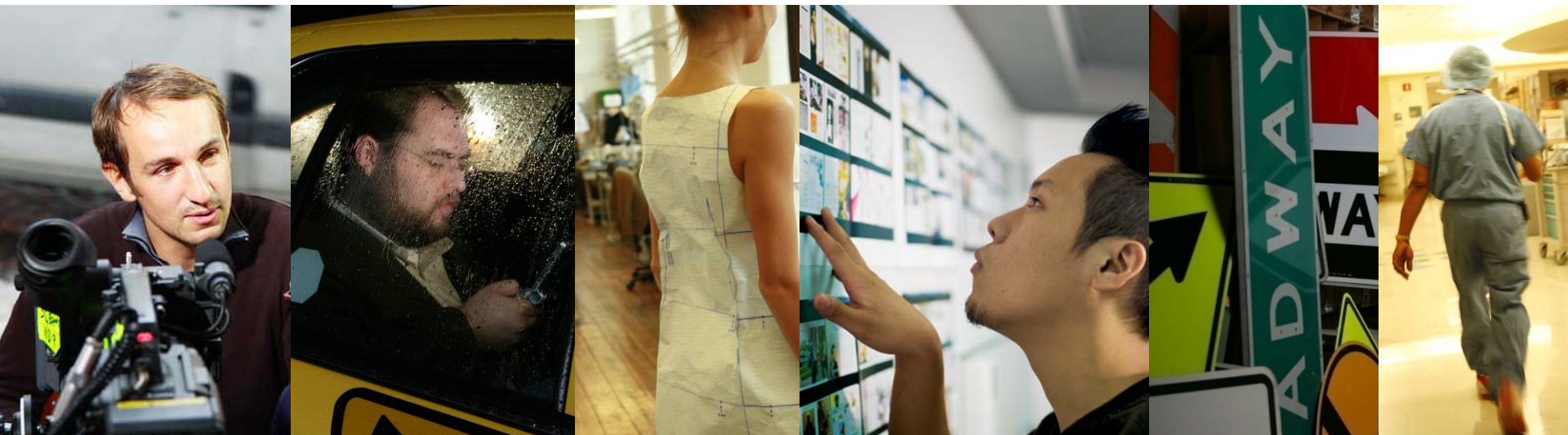
# Demo of Mars



# Summary

- Mars has feature parity with PDF
- Componentized model representing different aspects of a document
- Modular approach simplifies document creation/modification
- Decentralization means reuse is simpler
- Better integration with modern XML workflows

<http://www.adobe.com/go/mars>



**Revolutionizing**  
how the world engages  
with ideas and information

