



# Mathoid

Robust, Scalable, Fast and Accessible Math  
Rendering for Wikipedia

**Math on Wikipedia Meeting**

**29.5.2015 15h (CEST)**

#wikimedia-services on freenode

With slides from CICM 2014 by  
Moritz Schubotz and Gabriel Wicke

$$E = mc^2 \quad E = mc^2$$

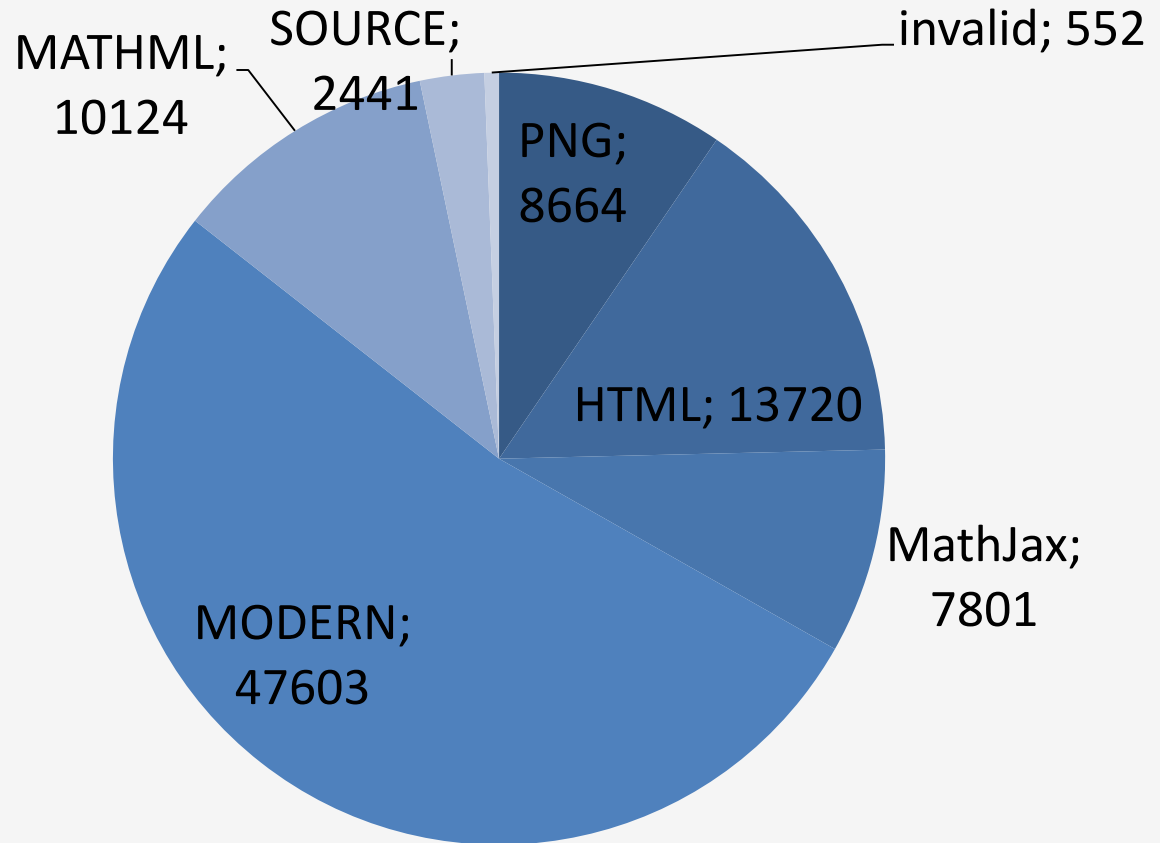


# History of Math and Wikipedia

- Math support since 2003
- 10-2010 Client side MathJax support
- 11-2011 MathML setting is removed
- 10-2013 Mathoid implementation ready
- 06-2014 Majority of the new code is reviewed
- 09-2014 Mathoid is deployed in production
- 03-2015 Mathoid 0.2.6 with speech output
- 05-2015 Mathoid 0.2.8 complies with MediaWiki services template



# The users preferences (2014)





# Bringing MathML to Wikipedia

- Dimensions
  - Coverage
  - Speed
  - Robustness
  - Maintainability
  - Accessibility

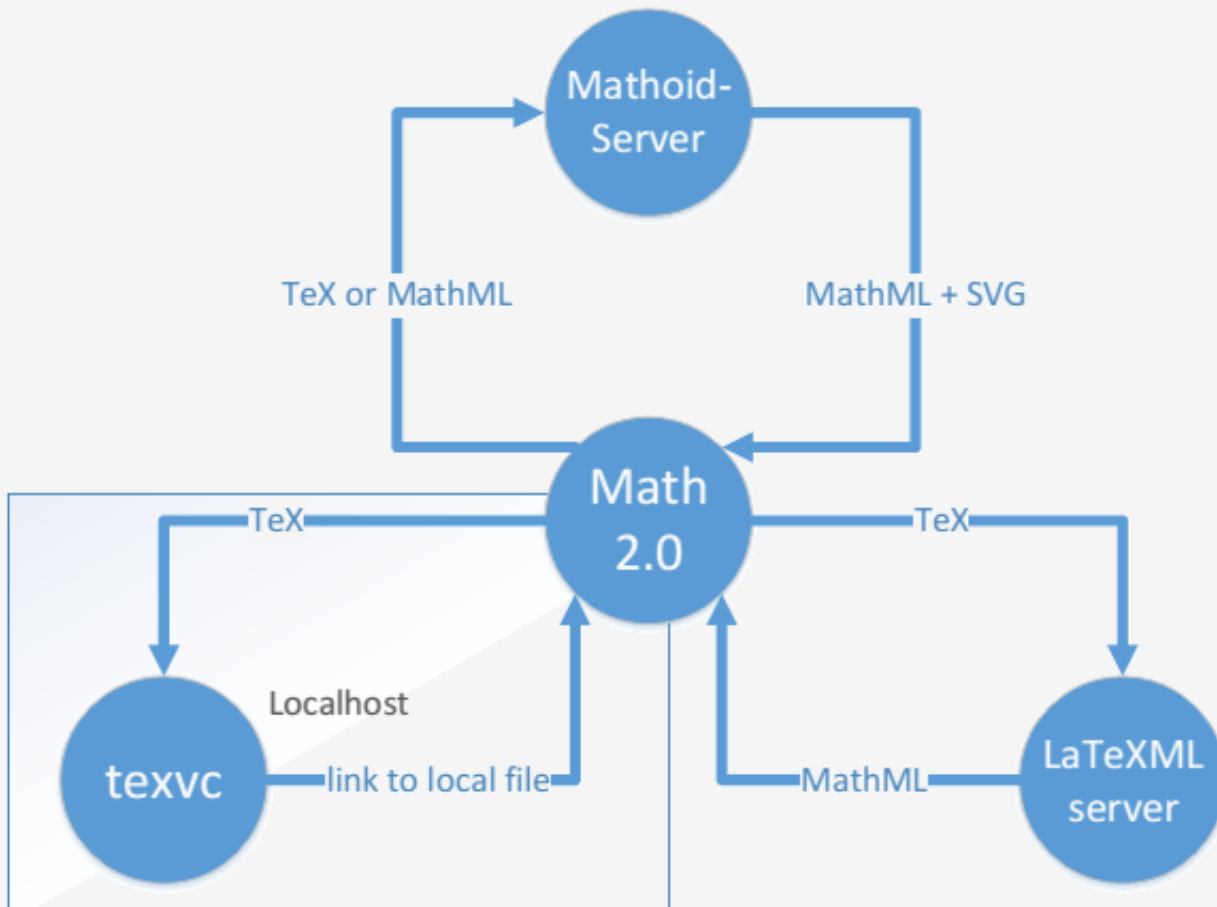


# Browser support

- MathML support in Firefox
- No MathML support in Chrome
- Fallback SVG images



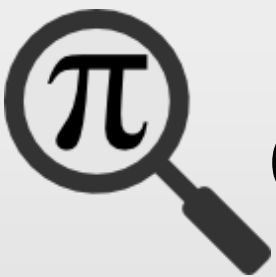
# Mathoid





# Wikipedia Dataset

- Wikipedia (en) 446 485 formulae (27 671 pages)
  - ~280k distinct formula
  - ~3GB formulae (presentation + content markup)
  - Generation on workstation 150MB source data



# Comparison of rendering methods

	texvc	L <sup>A</sup> T <sub>E</sub> XML	Mathoid
relative speed	1	0.3	5
image output	PNG	PNG	SVG
presentation MathML coverage	low	high	high
content MathML output	no	no	yes
webservice	no	yes	yes
approximate space required on webserver	1GB	0	0
language	OCaml	Perl	JavaScript
maintained by	nobody	NIST	MathJax

LaTeXML got faster and supports SVG now





# Mathoid needs you

- Help with testing of MathML in Wikipedia
- <http://math-preview.wmflabs.org>
- Report bugs  
<https://phabricator.wikimedia.org/tag/math/>
- [Review code](#)





# Technical Details: Caching

Rendering	Database Table	Page Cache	Browser cache
PNG	filename	Special page	yes
SVG	yes	Special page	yes
MathML	yes	Page output	
Livetime	infitiy	one week (Specialpage)	One hour (Specialpage)

Texvc generates an outputhash for each PNG image. This outputhash can not be computed from the TeX input without running texvc, and changes after texvc is recompiled. In the database the filename of the PNG file was stored. The specialpage gets the hash of the tex input to retrieve the image. If the image file is considered as correct (valid XML for SVG and existing for PNG) the special page header sets caching information.

Currently mathoid displays the texvc generated PNG for IE6 and older.



# Proposed changes (1)

- PNG generation independent of texvc
  - Currently JAVA
  - Next step Node SVG 2 PNG conversion
- Minimal change database caching layer

Rendering	Database Table	Page Cache	Browser cache
PNG	yes	Special page	yes
SVG	yes	Special page	yes
MathML	yes	Page output	
Livetime	infitiy	one week (Specialpage)	One hour (Specialpage)



## Proposed changes (2)

- Use standard Image pipeline for image fall back
- Minimal change database caching layer

Rendering	Database Table	Page Cache	Browser cache
MathML	yes	Page output	
Livetime	infitiy	one week (Specialpage)	One hour (Specialpage)



## Proposed changes (3)

- Completely remove database caching layer
- MathSearch extension maintains math index and is updated via the (already existing) hook