

Wikisource technical infrastructure

What we have done and what we could do

II

Thomas Pellissier Tanon

User:Tpt

@Tpt93

Wikisource Conference 2015



- Current state:
 - 4 millions of pages
 - 2.1 millions of proofread pages
 - 600 active editors (> 5 edits)

- Current state:
 - 4 millions of pages
 - 2.1 millions of proofread pages
 - 600 active editors (> 5 edits)
- Strong issues:
 - books not easily accessible
 - no real bibliographic database
 - contributing is quite difficult

- MediaWiki

- MediaWiki
- but with custom extensions like ProofreadPage

- MediaWiki
- but with custom extensions like ProofreadPage
- developed and maintained by volunteer contributors and a few GSoC projects

- 1 What we have done
- 2 What we could do

What we have done

- Is a "magic" export tool

Lettere à M. Dacier relative à l'alphabet des hiéroglyphes phonétiques

[66]

ture. Les cartouches d
méridionale renferm
TMHTIHNΣ (César I
cartouches placés ver
nale du même obélisque, torment la légende :
AOTKPΤAKHΣEPΣ TMHTIHNΣ EBETΣ.^[66] l'empereur César Domitien Auguste.

8° Le nom de Vespasien son Père se lit dans un des cartouches supérieurs de la même face, compris dans la formule idéographique qui a reçu la puissance venant de OYΣΠEHNΣ son père.^[67] les quatre premiers signes de ce cartouche sont trop rapprochés sur la gravure de Kircher.

9° Il existe dans la partie orientale de l'île de Philæ, un édifice fort élégant, mais dont la décoration hiéroglyphique n'a jamais été terminée. Du nombre des parties complètes, sont deux entrecolonnements dont l'un a été dessiné, dans tous ses détails, par la Commission d'Égypte.^[68] Les cartouches dont il est chargé, se rapportent tous à l'empereur Trajan. L'image en pied de ce bon prince, faisant une offrande à Isis et à Arouéris, est accompagnée de deux cartouches contenant les mots AOTKPTP KHΣEPΣ NPO TPHNΣ... (l'empereur César Nerva Trajan)^[69] ; et la légende TPHNΣ KHΣEPΣ (Trajan César) toujours vivant^[70], renfermée dans un

cartouche termine aussi la colonne perpendiculaire des hiéroglyphes sculptée à la droite du bas-relief. La base de ce même entrecolonnement est ornée de neuf petits cartouches. Celui du centre, un peu plus grand que les huit autres, soutenu par deux Uréus ou aspics royaux, renferme le nom de Trajan, TPHNΣ, avec l'épithète idéographique toujours vivant. Combiné avec celui de droite et celui de gauche, il produit la légende suivante : l'empereur toujours vivant ; Trajan toujours vivant ; César germe éternel d'Isis. Les trois cartouches rangés à la droite de ces derniers, produisent les mots Trajan toujours vivant, César, Germanicus, Dacicus, toujours vivant, Enfin, les trois cartouches de la gauche donnent la légende : NERVA TRAJAN toujours vivant, EMPEREUR CÉSAR toujours vivant, AUGUSTE^[71] toujours vivant chéri d'Isis. Le nom de Trajan se lit encore sur le grand temple d'Ombois ; deux cartouches dessinés dans les ruines de ce monument, forment en effet la série AOTOKPTP KHEA NAOA - TPHNΣ (l'empereur César-Nerva-Trajan), SURNOMMÉ KPMNHKE, THKKE (Germanicus, Dacicus)^[72] ; ce qui est encore, mot pour mot, la légende des médailles grecques de cet empereur frappées en Égypte.

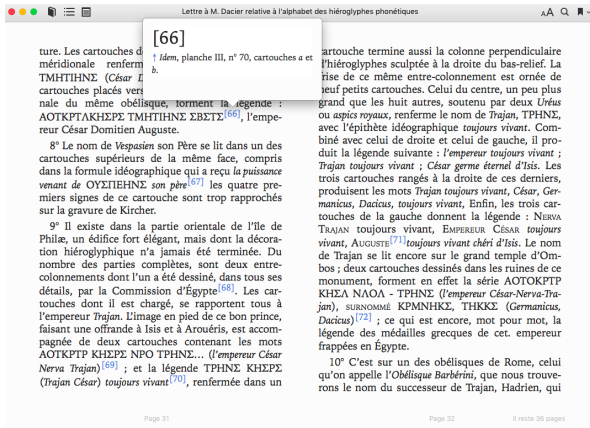
10° C'est sur un des obélisques de Rome, celui qu'on appelle l'Obélisque Barbérini, que nous trouverons le nom du successeur de Trajan, Hadrien, qui

Page 31

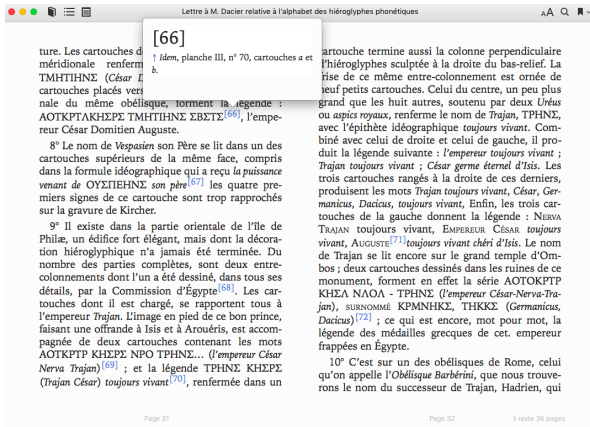
Page 32

Il reste 36 pages

- Is a "magic" export tool
- adapted to Wikisource needs

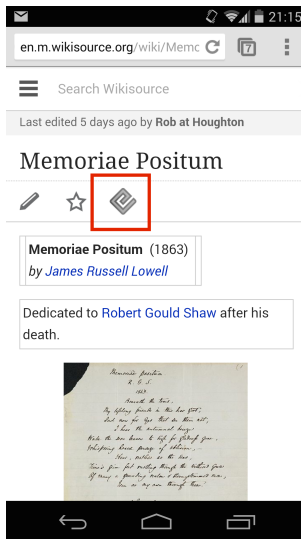


- Is a "magic" export tool
- adapted to Wikisource needs
- ePub is the base format

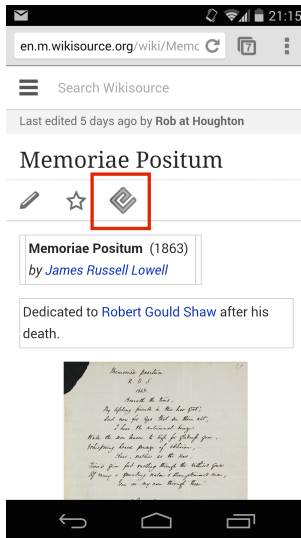


- Migrated on Wikimedia Tools labs at <https://tools.wmflabs.org/wsexport>

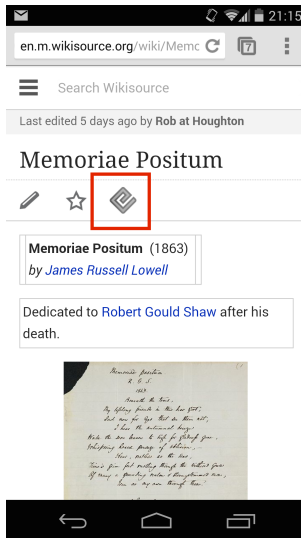
- Migrated on Wikimedia Tools labs at <https://tools.wmflabs.org/wsexport>
- Integrated in the UI of most of Wikisources



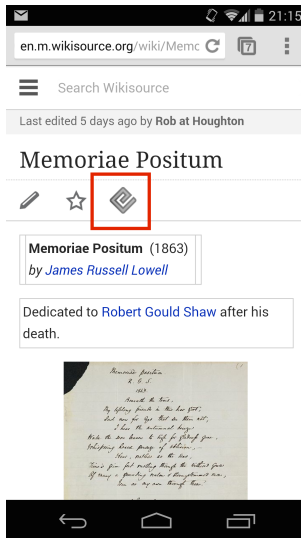
- Migrated on Wikimedia Tools labs at <https://tools.wmflabs.org/wsexport>
- Integrated in the UI of most of Wikisources
- 48,000 exports in October 2015



- Migrated on Wikimedia Tools labs at <https://tools.wmflabs.org/wsexport>
- Integrated in the UI of most of Wikisources
- 48,000 exports in October 2015
- Supports PDF, mobi...



- Migrated on Wikimedia Tools labs at <https://tools.wmflabs.org/wsexport>
- Integrated in the UI of most of Wikisources
- 48,000 exports in October 2015
- Supports PDF, mobi...
- ePub 3 is the default format



Refactoring of ProofreadPage

Goals:

- More maintainable code
- Use new MediaWiki features (ContentHandler...)

Refactoring of ProofreadPage

Goals:

- More maintainable code
- Use new MediaWiki features (ContentHandler...)
- Better performances
- Less breakages

Refactoring of ProofreadPage

Goals:

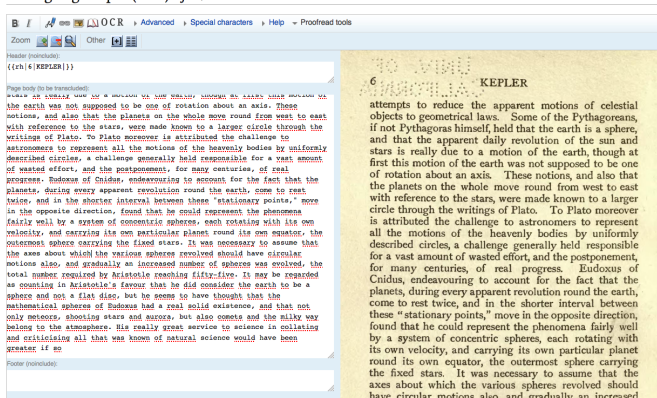
- More maintainable code
- Use new MediaWiki features (ContentHandler...)
- Better performances
- Less breakages
- Content validation

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP

Editing Page:Kepler(1920).djvu/12



The screenshot shows a web-based proofreading interface. At the top, there is a toolbar with icons for bold, italic, underline, and other text formatting, along with a search icon and a 'Proofread tools' link. Below the toolbar, there are sections for 'Header (not include):' and 'Footer (not include):', both containing empty curly braces. The main editing area is divided into two panes. The left pane, titled 'Page body (to be transcluded):', contains a large block of text with red squiggly lines indicating errors or corrections. The right pane shows a preview of the page, with the title 'KEPLER' and a paragraph of text. The text in the preview is: 'attempts to reduce the apparent motions of celestial objects to geometrical laws. Some of the Pythagoreans, if not Pythagoras himself, held that the earth is a sphere, and that the apparent daily revolution of the sun and stars is really due to a motion of the earth, though at first this motion of the earth was not supposed to be one of rotation about an axis. These notions, and also that the planets on the whole move round from west to east with reference to the stars, were made known to a larger circle through the writings of Plato. To Plato moreover is attributed the challenge to astronomers to represent all the motions of the heavenly bodies by uniformly described circles, a challenge generally held responsible for a vast amount of wasted effort, and the postponement, for many centuries, of real progress. Eudoxus of Cnidus, endeavouring to account for the fact that the planets, during every apparent revolution round the earth, come to rest twice, and in the shorter interval between these "stationary points," move in the opposite direction, found that he could represent the phenomena fairly well by a system of concentric spheres, each rotating with its own velocity, and carrying its own particular planet round its own equator, the outermost sphere carrying the fixed stars. It was necessary to assume that the axes about which the various spheres revolved should have circular motions also, and gradually an increased number of spheres was evolved, the total number required by Aristotle reaching fifty-five. It may be regarded as counting in Aristotle's favour that he did consider the earth to be a sphere and not a flat disc, but he seems to have thought that the mathematical spheres of Eudoxus had a real solid existence, and that not only meteors, shooting stars and auroras, but also comets and the milky way belong to the atmosphere. His really great service to science in collating and criticising all that was known of natural science would have been greater if so'.

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP
- Try to have not too badly architected code

Editing Page:Kepler(1920).djvu/12

The screenshot shows a proofreading interface for a page titled "6 KEPLER". The interface includes a header with navigation options like "Advanced", "Special characters", and "Help". Below the header, there is a text area containing the main body of the page, which is a historical text about Kepler's model of the universe. The text is displayed in a light blue background with a white border. The text describes Kepler's discovery that the planets move in elliptical orbits, a departure from the traditional circular orbits. It mentions that Kepler's model was a significant improvement over the geocentric model of Aristotle and Ptolemy, and that it was based on the principle of simplicity. The text is surrounded by a white border, and there are navigation controls on the right side of the page.

6 KEPLER

attempts to reduce the apparent motions of celestial objects to geometrical laws. Some of the Pythagoreans, if not Pythagoras himself, held that the earth is a sphere, and that the apparent daily revolution of the sun and stars is really due to a motion of the earth, though at first this motion of the earth was not supposed to be one of rotation about an axis. These notions, and also that the planets on the whole move round from west to east with reference to the stars, were made known to a larger circle through the writings of Plato. To Plato moreover is attributed the challenge to astronomers to represent all the motions of the heavenly bodies by uniformly described circles, a challenge generally held responsible for a vast amount of wasted effort, and the postponement, for many centuries, of real progress. Eudoxus of Cnidus, endeavouring to account for the fact that the planets, during every apparent revolution round the earth, come to rest twice, and in the shorter interval between these "stationary points," move in the opposite direction, found that he could represent the phenomena fairly well by a system of concentric spheres, each rotating with its own velocity, and carrying its own particular planet round its own equator, the outermost sphere carrying the fixed stars. It was necessary to assume that the axes about which the various spheres revolved should have circular motions also, and gradually an increased number of spheres was evolved, the total number required by Aristotle reaching fifty-five. It may be regarded as counting in Aristotle's favour that he did consider the earth to be a sphere and not a flat disc, but he seems to have thought that the mathematical spheres of Eudoxus had a real solid existence, and that not only meteors, shooting stars and auroras, but also comets and the milky way belong to the atmosphere. His really great service to science in collating and criticising all that was known of natural science would have been greater if so

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP
- Try to have not too badly architected code
- Automated tests

Editing Page:Kepler(1920).djvu/12

The screenshot shows a proofreading interface for a page titled "KEPLER". The interface includes a header with the page title, a toolbar with various editing tools, and a main text area. The text in the main area is a paragraph discussing the geocentric model of the universe and the heliocentric model proposed by Copernicus. The text is highlighted in blue, indicating it is selected for editing. The interface also includes a footer with the page title and a navigation bar at the bottom.

Header (noinclude):
{rh|6|KEPLER|}}

Page body (to be transcluded):
the earth was not supposed to be one of rotation about an axis. These motions, and also that the planets on the whole move round from west to east with reference to the stars, were made known to a larger circle through the writings of Plato. To Plato moreover is attributed the challenge to astronomers to represent all the motions of the heavenly bodies by uniformly described circles, a challenge generally held responsible for a vast amount of wasted effort, and the postponement, for many centuries, of real progress. Eudoxus of Cnidus, endeavouring to account for the fact that the planets, during every apparent revolution round the earth, come to rest twice, and in the shorter interval between these "stationary points," move in the opposite direction, found that he could represent the phenomena fairly well by a system of concentric spheres, each rotating with its own velocity, and carrying its own particular planet round its own equator, the outermost sphere carrying the fixed stars. It was necessary to assume that the axes about which the various spheres revolved should have circular motions also, and gradually an increased number of spheres was evolved, the total number required by Aristotle reaching fifty-five. It may be regarded as counting in Aristotle's favour that he did consider the earth to be a sphere and not a flat disc, but he seems to have thought that the mathematical spheres of Eudoxus had a real solid existence, and that not only meteors, shooting stars and auroras, but also comets and the milky way belong to the atmosphere. His really great service to science in collating and criticising all that was known of natural science would have been greater if so

Footer (noinclude):

Refactoring of ProofreadPage

Done:

- Rewrite editing interfaces in PHP
- Try to have not too badly architected code
- Automated tests
- JSON encoding of Page: pages in API

Editing Page:Kepler(1920).djvu/12

Header (noinclude):
{rh|6|KEPLER|}}

Page body (to be transcluded):
the earth was not supposed to be one of rotation about an axis. These motions, and also that the planets on the whole move round from west to east with reference to the stars, were made known to a larger circle through the writings of Plato. To Plato moreover is attributed the challenge to astronomers to represent all the motions of the heavenly bodies by uniformly described circles, a challenge generally held responsible for a vast amount of wasted effort, and the postponement, for many centuries, of real progress. Eudoxus of Cnidus, endeavouring to account for the fact that the planets, during every apparent revolution round the earth, come to rest twice, and in the shorter interval between these "stationary points," move in the opposite direction, found that he could represent the phenomena fairly well by a system of concentric spheres, each rotating with its own velocity, and carrying its own particular planet round its own equator, the outermost sphere carrying the fixed stars. It was necessary to assume that the axes about which the various spheres revolved should have circular motions also, and gradually an increased number of spheres was evolved, the total number required by Aristotle reaching fifty-five. It may be regarded as counting in Aristotle's favour that he did consider the earth to be a sphere and not a flat disc, but he seems to have thought that the mathematical spheres of Eudoxus had a real solid existence, and that not only meteors, shooting stars and auroras, but also comets and the milky way belong to the atmosphere. His really great service to science in collating and criticising all that was known of natural science would have been greater if so

Footer (noinclude):

6 KEPLER

attempts to reduce the apparent motions of celestial objects to geometrical laws. Some of the Pythagoreans, if not Pythagoras himself, held that the earth is a sphere, and that the apparent daily revolution of the sun and stars is really due to a motion of the earth, though at first this motion of the earth was not supposed to be one of rotation about an axis. These notions, and also that the planets on the whole move round from west to east with reference to the stars, were made known to a larger circle through the writings of Plato. To Plato moreover is attributed the challenge to astronomers to represent all the motions of the heavenly bodies by uniformly described circles, a challenge generally held responsible for a vast amount of wasted effort, and the postponement, for many centuries, of real progress. Eudoxus of Cnidus, endeavouring to account for the fact that the planets, during every apparent revolution round the earth, come to rest twice, and in the shorter interval between these "stationary points," move in the opposite direction, found that he could represent the phenomena fairly well by a system of concentric spheres, each rotating with its own velocity, and carrying its own particular planet round its own equator, the outermost sphere carrying the fixed stars. It was necessary to assume that the axes about which the various spheres revolved should have circular motions also, and gradually an increased

- Download/print
 - Download as PDF
 - Download as EPUB
 - Printable version
 - Create a book
- Other projects**
 - Wikimedia Commons
 - Meta-Wiki
 - Wikispecies
 - Wikibooks
 - Wikidata
 - Wikinews
 - Wikipedia
 - Wikiquote
 - Wikivoyage
- In other languages 
 - Ænglisc
 - العربية
 - অসমীয়া
 - Azərbaycanca
 - Беларуская
 - Български

is accompanied by detailed photographs
can read cuneiform.

The ruins of Susa now being excavated by already yielded important results. He was from inscriptions found in the ruins of Bab important monuments of the Babylonian k Elamite kings, to their capital, Susa. Wher recovering from the ruins of Susa these m before he found the stele of Narâm-Sin c. Tigro-Euphrates valley nearly 6,000 years important of which is the stele of Hammur

This code is the oldest collection of public conditions existing in Babylonia 4,000 yea principles on which our social legislation is

(Read on... November's featured text

or see [all featured texts](#).

CUR

The current **Community collaboration**

Recent collaborations: [Tenth Anniversary C](#)

An Internet Archive to Commons import tool (<http://tools.wmflabs.org/ia-upload>)

Upload books from Internet Archive.

Upload a file

Id of the file in Internet Archive:

champollionsein02hartuoft

File name on Commons:

Hartleben_-_Champollion_sein_Leben_und_sein_Werk,_1906_band_2.djvu

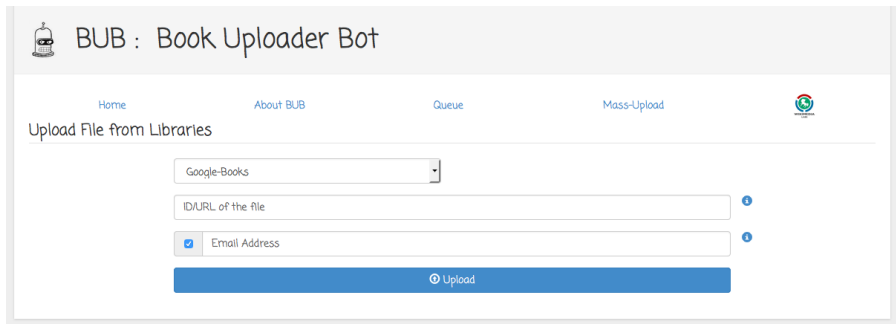
Commons description:

```
== {{int:filedesc}} ==
{{Book
|Author    = Hartleben, Hermine
|Editor    =
|Translator =
|Illustrator =
|Title     = Champollion : sein Leben und sein Werk
|Subtitle  =
|Series title =
|Volume    = 02
|Edition   =
|Publisher = Berlin : Weidmann
|Printer   =
|Date      = 1906
|City      =
|Language  = {{language|de}}
}}
```

- The author "Hartleben, Hermine" doesn't have a [creator](#) template. Isn't he known under an other name or do you want to [create it](#) ?

Upload (Warning: the upload may take more than one minute !)

An import tool for Internet Archive from Google Books and other sources
(<https://tools.wmflabs.org/bub/>)



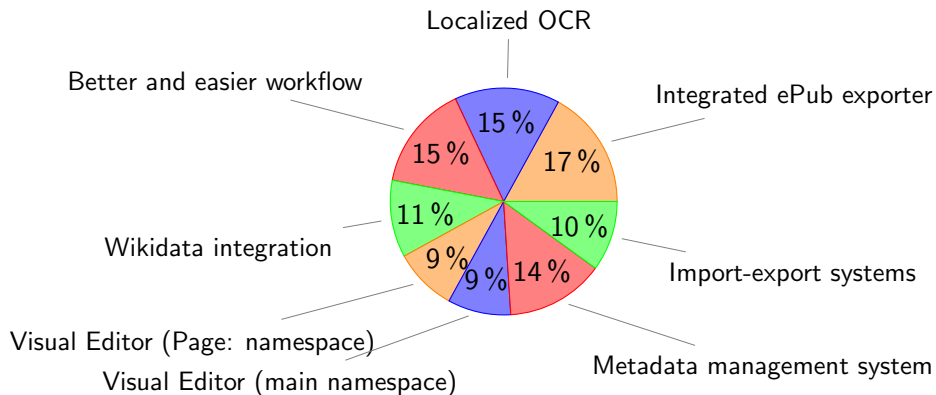
The screenshot shows the web interface for BUB: Book Uploader Bot. At the top left is a small robot icon. The title "BUB : Book Uploader Bot" is displayed in a large, handwritten-style font. Below the title is a navigation bar with four links: "Home", "About BUB", "Queue", and "Mass-Upload". On the right side of the navigation bar is a small circular logo. The main content area is titled "Upload File from Libraries". It features a dropdown menu currently set to "Google-Books". Below this are two input fields: "ID/URL of the file" and "Email Address", each with a blue information icon to its right. The "Email Address" field has a checked checkbox on its left. At the bottom of the form is a large blue button with a white circular icon containing an upward-pointing arrow and the text "Upload".

What we could do

Some ideas...

A Wikisource contributors survey done in Fall 2013 with 251 answers

Figure: What do you think are the core priorities for the Wikisource community?




- Improve Parsoid rendering of Wikisource content

- Improve Parsoid rendering of Wikisource content
- Support tags used on Wikisource (pages, poem, section...)

VisualEditor Image support

- Improve Parsoid rendering of Wikisource content
- Support tags used on Wikisource (pages, poem, section...)
- Custom interface for Page: pages



The screenshot shows the VisualEditor interface for a Wikisource page. The browser address bar shows the page title: "Page:La Brière - Champollion inconnu.djvu/5". The interface includes a top navigation bar with "Page", "Discussion", and "Image" tabs, and a right-hand navigation bar with "Read", "Edit", "Edit source", "View history", and "More" options. A search bar is also present. The main editing area displays the page title and a green validation message: "This page has been validated". The content is split into two columns. The left column shows the source code for the first section, including the heading "CHAMPOLLION ENFANT" and the start of a paragraph. The right column shows the rendered HTML output, including the heading "CHAMPOLLiON ENFANT" and the rendered text of the paragraph. The text in the rendered output is: "Le génial savant qui, par une divinatrice intuition, découvrit, au commencement de ce siècle, la clé perdue de l'écriture égyptienne, déchirant soudain les voiles d'une histoire mystérieuse, exhumant des entrailles du passé toute une science nouvelle, laquelle dépassa presque aussitôt les balbutiants essais, et atteint déjà de notre temps sa maturité pleine, Jean-François Champollion, l'une des gloires les plus pures et les plus universellement reconnues de notre panthéon national, n'est inconnu d'aucun Français. Les annales scientifiques ont salué sa courte vie de quarante-deux ans, et son œuvre prodigieuse. S'il restait encore, dans ses admirables travaux quelque mérite à célébrer, ce serait aux initiés, professionnels de l'Égyptologie, qu'il appartiendrait de compléter son panégyrique." Below the text, there is a note: "There are no references on this page to include in this list."

- Custom edit interface for Page: and Index: pages

← Editing Page:La Brière - Ch... Next

```
<noinclude><pagequality level="4"
user="Maniak" /><div class="pagetext">
```

</noinclude>{{petit titre|}}

{{petit titre|CHAMPOLLION ENFANT}}

Le génial savant qui, par une divinatrice intuition, découvrit, au commencement de ce siècle, la clé perdue de l'écriture égyptienne, déchirant soudain les voiles d'une histoire mystérieuse, exhumant des entrailles du passé

Mobile support

- Custom edit interface for Page: and Index: pages
- We should have a nice UI for both browsing and editing

← Editing Page:La Brière - Ch... Next

```
<noinclude><pagequality level="4"
user="Maniak" /><div class="pagetext">
```

</noinclude>{{petit titre|}}

{{petit titre|CHAMPOLLION ENFANT}}

Le génial savant qui, par une divinatrice intuition, découvrit, au commencement de ce siècle, la clé perdue de l'écriture égyptienne, déchirant soudain les voiles d'une histoire mystérieuse, exhumant des entrailles du passé

- Custom edit interface for Page: and Index: pages
- We should have a nice UI for both browsing and editing
- Future: gamification?

← Editing Page:La Brière - Ch... Next

```
<noinclude><pagequality level="4"
user="Maniak" /><div class="pagetext">
```

</noinclude>{{petit titre|}}

{{petit titre|CHAMPOLLION ENFANT}}

Le génial savant qui, par une divinatrice intuition, découvrit, au commencement de ce siècle, la clé perdue de l'écriture égyptienne, déchirant soudain les voiles d'une histoire mystérieuse, exhumant des entrailles du passé

- Improve Wsexport performances

Exportation of content

- Improve Wsexport performances
- Use Parsoid

Exportation of content

- Improve Wsexport performances
- Use Parsoid
- Nice book browsing interface + OPDS?

- Improve Wsexport performances
- Use Parsoid
- Nice book browsing interface + OPDS?
- Integrated inside of Wikimedia infrastructure?

Some other ideas

- Use Wikidata as much as possible

Some other ideas

- Use Wikidata as much as possible
- Gamification (capcha...)

Some other ideas

- Use Wikidata as much as possible
- Gamification (capcha...)
- ...

What we need now:

- Stronger interwiki collaboration

What we need now:

- Stronger interwiki collaboration
- People to build the Wikisource of tomorrow

What we need now:

- Stronger interwiki collaboration
- People to build the Wikisource of tomorrow
- Stronger support from the Wikimedia movement

Thanks a lot for your attention!

