Challenges for Digital Literacy in the Humanities

The Open, Community-Based and Multilinguistic Approach of The Programming Historian

https://programminghistorian.org / @ProgHist

Sofia Papastamkou, CNRS/U Lille, France
Jessica Parr, Simmons College, MA, USA
Riva Quiroga, Universidad Católica de Chile, Chile

Introduction

Timeline of Programming Historian

2008	2012	2017	2019	2021
Programming Historian 1	**Programming Historian (2)**	**Programming Historian en español**	**Programming Historian en français**	**Programming Historian em português**
William J. Turkel, Alan McEachern	Editorial Board	ES team (2016)	FR team (2018)	PT team (2020)

Programming Historian at a glance

- **Diamond OA** journal (CC BY, no APC)
- **Run by 28 volunteer editors** as to March 2021 (46 from 2008 to present)
- Content: tutorials on digital methods, tools, workflows for historians and humanists
- **4 linguistic versions**: EN (2012), ES (2017), FR (2019), PT (2021)
- Tutorials: 84 (EN - original), 49 (ES - 6 original), 15 (FR), 1 (PT)

Programming Historian: Core Elements

1. **Open Values**
2. **Community-based/run**
3. **Multilingualism that accounts for linguistic diversity within language communities.**

Open Values

Communication and Infrastructure

Our communication happens in synchronous and asynchronous forms.

Asynchronously:

- Google email list
- Discussion over Github

Synchronously:

- Quarterly team meetings (i.e. Global meets every 3 months).
- Full project team meetings every other month.

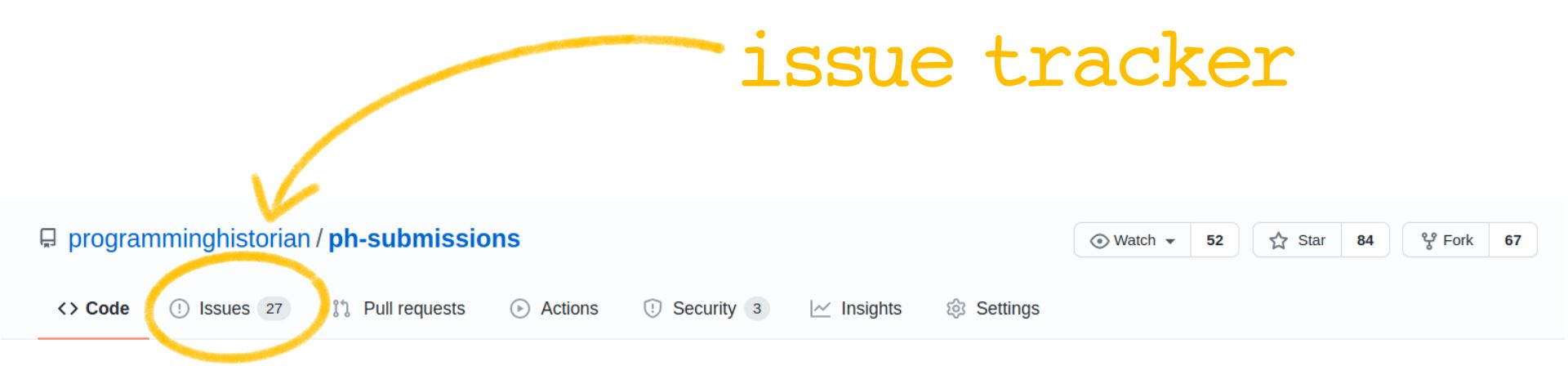
Where the world builds software

Millions of developers and companies build, ship, and maintain their software on GitHub—the largest and most advanced development platform in the world.

Email

Sign up for GitHub

issue tracker



programminghistorian / **ph-submissions**

Watch 52 | Star 84 | Fork 67

Code | Issues 27 | Pull requests | Actions | Security 3 | Insights | Settings

gh-pages | **16** branches | **0** tags | Go to file | Add file | Code

jenniferisasi Update analisis-de-sentimientos-r.md — 753bd34 12 hours ago — **3,977** commits

.github	Update ISSUE_TEMPLATE	28 days ago
.sass-cache	More metadata and adding avatarg	13 months ago
_data	Last proofing	16 months ago
_includes	corrects js error	2 years ago
_layouts	add big warning to lesson pages	2 years ago
assets	Add files via upload	19 days ago
css	Remove line numbers from general CSS file	5 years ago
en	Update georeferencing-and-displaying-historical-maps.md	8 days ago

About

A work-in-progress site for submitting and reviewing new Programming Historian lessons

programminghistorian.github.io/ph...

Readme

Releases

No releases published
Create a new release

Packages

Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS #349

New issue

Open **sferna109** opened this issue 7 days ago · 0 comments

sferna109 commented 7 days ago — Contributor

The Programming Historian has received the following tutorial on 'Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS' by **@ericayhayes**. This lesson is now under review and can be read at:

http://programminghistorian.github.io/ph-submissions/en/lessons/georeferencing-and-displaying-historical-maps

Please feel free to use the line numbers provided on the preview if that helps with anchoring your comments, although you can structure your review as you see fit.

I will act as editor for the review process. My role is to solicit two reviews from the community and to manage the discussions, which should be held here on this forum. I have already read through the lesson and provided feedback, to which the author has responded.

Members of the wider community are also invited to offer constructive feedback which should post to this message thread, but they are asked to first read our Reviewer Guidelines (http://programminghistorian.org/reviewer-guidelines) and to adhere to our anti-harassment policy (below). We ask that all reviews stop after the second formal review has been submitted so that the author can focus on any revisions. I will make an announcement on this thread when that has occurred.

I will endeavor to keep the conversation open here on Github. If anyone feels the need to discuss anything privately, you are welcome to email me.

Our dedicated Ombudsperson is (Ian Milligan - http://programminghistorian.org/en/project-team). Please feel free to contact him at any time if you have concerns that you would like addressed by an impartial observer. Contacting the ombudsperson will have no impact on the outcome of any peer review.

Anti-Harassment Policy

This is a statement of the Programming Historian's principles and sets expectations for the tone and style of all correspondence between reviewers, authors, editors, and contributors to our public forums.

> The Programming Historian is dedicated to providing an open scholarly environment that offers community participants the freedom to thoroughly scrutinize ideas, to ask questions, make suggestions, or to requests for clarification, but also provides a harassment-free space for all contributors to the project, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion, or technical experience. We do not tolerate harassment or ad hominem attacks of community participants in any form. Participants violating these rules may be expelled from the community at the discretion of the editorial board. Thank you for helping us to create a safe space.

Assignees
sferna109

Labels
English submission

Milestone
No milestone

Linked pull requests
Successfully merging a pull request may close this issue.
None yet

1 participant

Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS #349

New issue

Open **sferna109** opened this issue 7 days ago · 0 comments

sferna109 commented 7 days ago — Contributor

The Programming Historian has received the following tutorial on 'Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS' by **@ericayhayes**. This lesson is now under review and can be read at:

http://programminghistorian.github.io/ph-submissions/en/lessons/georeferencing-and-displaying-historical-maps

Please feel free to use the line numbers provided on the preview if that helps with anchoring your comments, although you can structure your review as you see fit.

I will act as editor for the review process. My role is to solicit two reviews from the community and to manage the discussions, which should be held here on this forum. I have already read through the lesson and provided feedback, to which the author has responded.

Members of the wider community are also invited to offer constructive feedback which should post to this message thread, but they are asked to first read our Reviewer Guidelines (http://programminghistorian.org/reviewer-guidelines) and to adhere to our anti-harassment policy (below). We ask that all reviews stop after the second formal review has been submitted so that the author can focus on any revisions. I will make an announcement on this thread when that has occurred.

I will endeavor to keep the conversation open here on Github. If anyone feels the need to discuss anything privately, you are welcome to email me.

Our dedicated Ombudsperson is (Ian Milligan - http://programminghistorian.org/en/project-team). Please feel free to contact him at any time if you have concerns that you would like addressed by an impartial observer. Contacting the ombudsperson will have no impact on the outcome of any peer review.

Anti-Harassment Policy

This is a statement of the Programming Historian's principles and sets expectations for the tone and style of all correspondence between reviewers, authors, editors, and contributors to our public forums.

> The Programming Historian is dedicated to providing an open scholarly environment that offers community participants the freedom to thoroughly scrutinize ideas, to ask questions, make suggestions, or to requests for clarification, but also provides a harassment-free space for all contributors to the project, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion, or technical experience. We do not tolerate harassment or ad hominem attacks of community participants in any form. Participants violating these rules may be expelled from the community at the discretion of the editorial board. Thank you for helping us to create a safe space.

Assignees
sferna109

Labels
English submission

Milestone
No milestone

Linked pull requests
Successfully merging a pull request may close this issue.
None yet

1 participant

Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS #349

New issue

Open **sferna109** opened this issue 7 days ago · 0 comments

sferna109 commented 7 days ago — Contributor

The Programming Historian has received the following tutorial on 'Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS' by **@ericayhayes**. This lesson is now under review and can be read at:

http://programminghistorian.github.io/ph-submissions/en/lessons/georeferencing-and-displaying-historical-maps

Please feel free to use the line numbers provided on the preview if that helps with anchoring your comments, although you can structure your review as you see fit.

I will act as editor for the review process. My role is to solicit two reviews from the community and to manage the discussions, which should be held here on this forum. I have already read through the lesson and provided feedback, to which the author has responded.

Members of the wider community are also invited to offer constructive feedback which should post to this message thread, but they are asked to first read our Reviewer Guidelines (http://programminghistorian.org/reviewer-guidelines) and to adhere to our anti-harassment policy (below). We ask that all reviews stop after the second formal review has been submitted so that the author can focus on any revisions. I will make an announcement on this thread when that has occurred.

I will endeavor to keep the conversation open here on Github. If anyone feels the need to discuss anything privately, you are welcome to email me.

Our dedicated Ombudsperson is (Ian Milligan - http://programminghistorian.org/en/project-team). Please feel free to contact him at any time if you have concerns that you would like addressed by an impartial observer. Contacting the ombudsperson will have no impact on the outcome of any peer review.

Anti-Harassment Policy

This is a statement of the Programming Historian's principles and sets expectations for the tone and style of all correspondence between reviewers, authors, editors, and contributors to our public forums.

> The Programming Historian is dedicated to providing an open scholarly environment that offers community participants the freedom to thoroughly scrutinize ideas, to ask questions, make suggestions, or to requests for clarification, but also provides a harassment-free space for all contributors to the project, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion, or technical experience. We do not tolerate harassment or ad hominem attacks of community participants in any form. Participants violating these rules may be expelled from the community at the discretion of the editorial board. Thank you for helping us to create a safe space.

Assignees
sferna109

Labels
English submission

Milestone
No milestone

Linked pull requests
Successfully merging a pull request may close this issue.
None yet

1 participant

Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS #349

New issue

Open **sferna109** opened this issue 7 days ago · 0 comments

sferna109 commented 7 days ago — Contributor

The Programming Historian has received the following tutorial on 'Georeferencing and Displaying Historical Maps using Map Warper and StoryMap JS' by **@ericayhayes**. This lesson is now under review and can be read at:

http://programminghistorian.github.io/ph-submissions/en/lessons/georeferencing-and-displaying-historical-maps

Please feel free to use the line numbers provided on the preview if that helps with anchoring your comments, although you can structure your review as you see fit.

I will act as editor for the review process. My role is to solicit two reviews from the community and to manage the discussions, which should be held here on this forum. I have already read through the lesson and provided feedback, to which the author has responded.

Members of the wider community are also invited to offer constructive feedback which should post to this message thread, but they are asked to first read our Reviewer Guidelines (http://programminghistorian.org/reviewer-guidelines) and to adhere to our anti-harassment policy (below). We ask that all reviews stop after the second formal review has been submitted so that the author can focus on any revisions. I will make an announcement on this thread when that has occurred.

I will endeavor to keep the conversation open here on Github. If anyone feels the need to discuss anything privately, you are welcome to email me.

Our dedicated Ombudsperson is (Ian Milligan - http://programminghistorian.org/en/project-team). Please feel free to contact him at any time if you have concerns that you would like addressed by an impartial observer. Contacting the ombudsperson will have no impact on the outcome of any peer review.

Anti-Harassment Policy

This is a statement of the Programming Historian's principles and sets expectations for the tone and style of all correspondence between reviewers, authors, editors, and contributors to our public forums.

> The Programming Historian is dedicated to providing an open scholarly environment that offers community participants the freedom to thoroughly scrutinize ideas, to ask questions, make suggestions, or to requests for clarification, but also provides a harassment-free space for all contributors to the project, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion, or technical experience. We do not tolerate harassment or ad hominem attacks of community participants in any form. Participants violating these rules may be expelled from the community at the discretion of the editorial board. Thank you for helping us to create a safe space.

Assignees
sferna109

Labels
English submission

Milestone
No milestone

Linked pull requests
Successfully merging a pull request may close this issue.
None yet

1 participant

Reviewer Guidelines

Reviewing for the *Programming Historian* is a great way to learn new technical skills and engage with the digital humanities community. We go out of our way to make sure our reviewers get credit and recognition for their work. Because reviewers directly contribute to significantly improving lessons, you can take pride in that your work helps thousands of readers.

These guidelines are to help reviewers understand their role in the editorial process and to answer common questions about how to be most efficient and effective with your reviews.

cneud commented on 14 Jul 2020 • edited ▾ ...

Thank you **@akhlaghiandrew** for this submission. It encapsulates the complexity of OCR and MT into a suitable scope for digital humanities use. At first, I thought it may be too much to include OCR and MT in a single tutorial, but the way it is presented works for me and highlights the fact that OCR is often just a first stepping stone for working with digitized texts. I think it is also very helpful that there is good context about image quality and its implications on OCR/MT quality, despite the fact that image processing takes up a lot more space in the tutorial than the actual OCR.

Remarks, suggestions

Under **Acquire the data**, maybe mention that OCR is only suitable for detecting and recognizing printed/typewritten letters and not the handwritten annotations also present on the example pages (for which there are other solutions, e.g. Transkribus).

Under **OCR**: Tesseract is **one of the best** (there are other, competitive open source OCR engines, e.g. Calamari, OCRopus). Google no longer maintains Tesseract btw! Ray Smith (employed by Google) started it originally at HP and developed it further at Google, but currently maintenance is being done by community members rather than Ray/Google. Furthermore, even training Tesseract for calligraphy or handwritten text would likely not work, since there are other segmentation methods required too, for non-printed glyphs. Therefore maybe better mention "unique scripts like Blackletter asf.".

Following on, in L76, one might want to explain that the param `-l` is for specifying the language to use for recognition. You might also want to mention that there are additional command line params (https://github.com/tesseract-ocr/tesseract/wiki#running-tesseract) e.g. to output different formats that also include structural tagging in addition to plain text (ALTO, hOCR).

Generally, I would suggest adding a link to the documentation for Tesseract (https://tesseract-ocr.github.io/).

Spelling, phrasing

L28: learn how **to** make scripts
L34: The linked Programming Historian tutorial contains instructions for ~~how~~ Windows users, **who** can use an emulator to use BASH.
L36: Open **a** Terminal
L40: I selected these documents for two reason**s**
L40: One, the documents are high quality scans, but have ~~errors~~ **defects** common to many archival documents
L40: so we will be able to judge the ~~fidelity~~ **quality/accuracy** of our machine translations.
L45: That is, the **text lines in the** document should appear straight in the picture.
L50: Mac Instal**l**ation
L53: On MacOS, the instal**l**ation requires two simple commands
L57: Windows Instal**l**ation
L61: OCR programs will only ~~take~~ **accept** image files (JPG, TIFF, PNG) **as input**
L65: Better use `ppi` (pixels-per-inch) as `dpi` (dots-per-inch) is only relevant to printing (cf. https://photographylife.com/dpi-vs-ppi)
L65: this ~~line of commands~~ **command** converts (it is a single command, despite multiple params ;)
L67: For example, they may be ~~a~~ skew or ~~have~~ uneven brightness.
L67: For an idea of other ImageMagick options that can ~~help~~ **improve** OCR **quality**
L70: Tesseract 4.1
L70: it might be worth training your own OCR ~~algorithm~~ **model**
L70: "typewritten" instead of "type written"
L70: and ~~correctly~~ **correctly** identify imperfect letters
L74: Windows instal**l**ation
L76: Our output is a plain text file in ~~English~~ **Russian**.

cneud commented on 14 Jul 2020 • edited

Thank you **@akhlaghiandrew** for this submission. It encapsulates the complexity of OCR and MT into a suitable scope for digital humanities use. At first, I thought it may be too much to include OCR and MT in a single tutorial, but the way it is presented works for me and highlights the fact that OCR is often just a first stepping stone for working with digitized texts. I think it is also very helpful that there is good context about image quality and its implications on OCR/MT quality, despite the fact that image processing takes up a lot more space in the tutorial than the actual OCR.

Remarks, suggestions

Under **Acquire the data**, maybe mention that OCR is only suitable for detecting and recognizing printed/typewritten letters and not the handwritten annotations also present on the example pages (for which there are other solutions, e.g. Transkribus).

Under **OCR**: Tesseract is **one of the best** (there are other, competitive open source OCR engines, e.g. Calamari, OCRopus). Google no longer maintains Tesseract btw! Ray Smith (employed by Google) started it originally at HP and developed it further at Google, but currently maintenance is being done by community members rather than Ray/Google. Furthermore, even training Tesseract for calligraphy or handwritten text would likely not work, since there are other segmentation methods required too, for non-printed glyphs. Therefore maybe better mention "unique scripts like Blackletter asf.".

Following on, in L76, one might want to explain that the param `-l` is for specifying the language to use for recognition. You might also want to mention that there are additional command line params (https://github.com/tesseract-ocr/tesseract/wiki#running-tesseract) e.g. to output different formats that also include structural tagging in addition to plain text (ALTO, hOCR).

Generally, I would suggest adding a link to the documentation for Tesseract (https://tesseract-ocr.github.io/).

Spelling, phrasing

L28: learn how **to** make scripts
L34: The linked Programming Historian tutorial contains instructions for ~~how~~ Windows users, **who** can use an emulator to use BASH.
L36: Open **a** Terminal
L40: I selected these documents for two reason**s**
L40: One, the documents are high quality scans, but have ~~errors~~ **defects** common to many archival documents
L40: so we will be able to judge the ~~fidelity~~ **quality/accuracy** of our machine translations.
L45: That is, the **text lines in the** document should appear straight in the picture.
L50: Mac Instal**l**ation
L53: On MacOS, the instal**l**ation requires two simple commands
L57: Windows Instal**l**ation
L61: OCR programs will only ~~take~~ **accept** image files (JPG, TIFF, PNG) **as input**
L65: Better use `ppi` (pixels-per-inch) as `dpi` (dots-per-inch) is only relevant to printing (cf. https://photographylife.com/dpi-vs-ppi)
L65: this ~~line of commands~~ **command** converts (it is a single command, despite multiple params ;)
L67: For example, they may be ~~a~~ skew or ~~have~~ uneven brightness.
L67: For an idea of other ImageMagick options that can ~~help~~ **improve** OCR **quality**
L70: Tesseract 4.1
L70: it might be worth training your own OCR ~~algorithm~~ **model**
L70: "typewritten" instead of "type written"
L70: and ~~correctly~~ **correctly** identify imperfect letters
L74: Windows instal**l**ation
L76: Our output is a plain text file in ~~English~~ **Russian**.

cneud commented on 14 Jul 2020 • edited

Thank you **@akhlaghiandrew** for this submission. It encapsulates the complexity of OCR and MT into a suitable scope for digital humanities use. At first, I thought it may be too much to include OCR and MT in a single tutorial, but the way it is presented works for me and highlights the fact that OCR is often just a first stepping stone for working with digitized texts. I think it is also very helpful that there is good context about image quality and its implications on OCR/MT quality, despite the fact that image processing takes up a lot more space in the tutorial than the actual OCR.

Remarks, suggestions

Under **Acquire the data**, maybe mention that OCR is only suitable for detecting and recognizing printed/typewritten letters and not the handwritten annotations also present on the example pages (for which there are other solutions, e.g. Transkribus).

Under **OCR**: Tesseract is **one of the best** (there are other, competitive open source OCR engines, e.g. Calamari, OCRopus). Google no longer maintains Tesseract btw! Ray Smith (employed by Google) started it originally at HP and developed it further at Google, but currently maintenance is being done by community members rather than Ray/Google. Furthermore, even training Tesseract for calligraphy or handwritten text would likely not work, since there are other segmentation methods required too, for non-printed glyphs. Therefore maybe better mention "unique scripts like Blackletter asf.".

Following on, in L76, one might want to explain that the param `-l` is for specifying the language to use for recognition. You might also want to mention that there are additional command line params (https://github.com/tesseract-ocr/tesseract/wiki#running-tesseract) e.g. to output different formats that also include structural tagging in addition to plain text (ALTO, hOCR).

Generally, I would suggest adding a link to the documentation for Tesseract (https://tesseract-ocr.github.io/).

Spelling, phrasing

L28: learn how **to** make scripts
L34: The linked Programming Historian tutorial contains instructions for ~~how~~ Windows users, **who** can use an emulator to use BASH.
L36: Open **a** Terminal
L40: I selected these documents for two reason**s**
L40: One, the documents are high quality scans, but have ~~errors~~ **defects** common to many archival documents
L40: so we will be able to judge the ~~fidelity~~ **quality/accuracy** of our machine translations.
L45: That is, the **text lines in the** document should appear straight in the picture.
L50: Mac Instal**l**ation
L53: On MacOS, the instal**l**ation requires two simple commands
L57: Windows Instal**l**ation
L61: OCR programs will only ~~take~~ **accept** image files (JPG, TIFF, PNG) **as input**
L65: Better use `ppi` (pixels-per-inch) as `dpi` (dots-per-inch) is only relevant to printing (cf. https://photographylife.com/dpi-vs-ppi)
L65: this ~~line of commands~~ **command** converts (it is a single command, despite multiple params ;)
L67: For example, they may be ~~a~~ skew or ~~have~~ uneven brightness.
L67: For an idea of other ImageMagick options that can ~~help~~ **improve** OCR **quality**
L70: Tesseract 4.1
L70: it might be worth training your own OCR ~~algorithm~~ **model**
L70: "typewritten" instead of "type written"
L70: and ~~correctly~~ **correctly** identify imperfect letters
L74: Windows instal**l**ation
L76: Our output is a plain text file in ~~English~~ **Russian**.

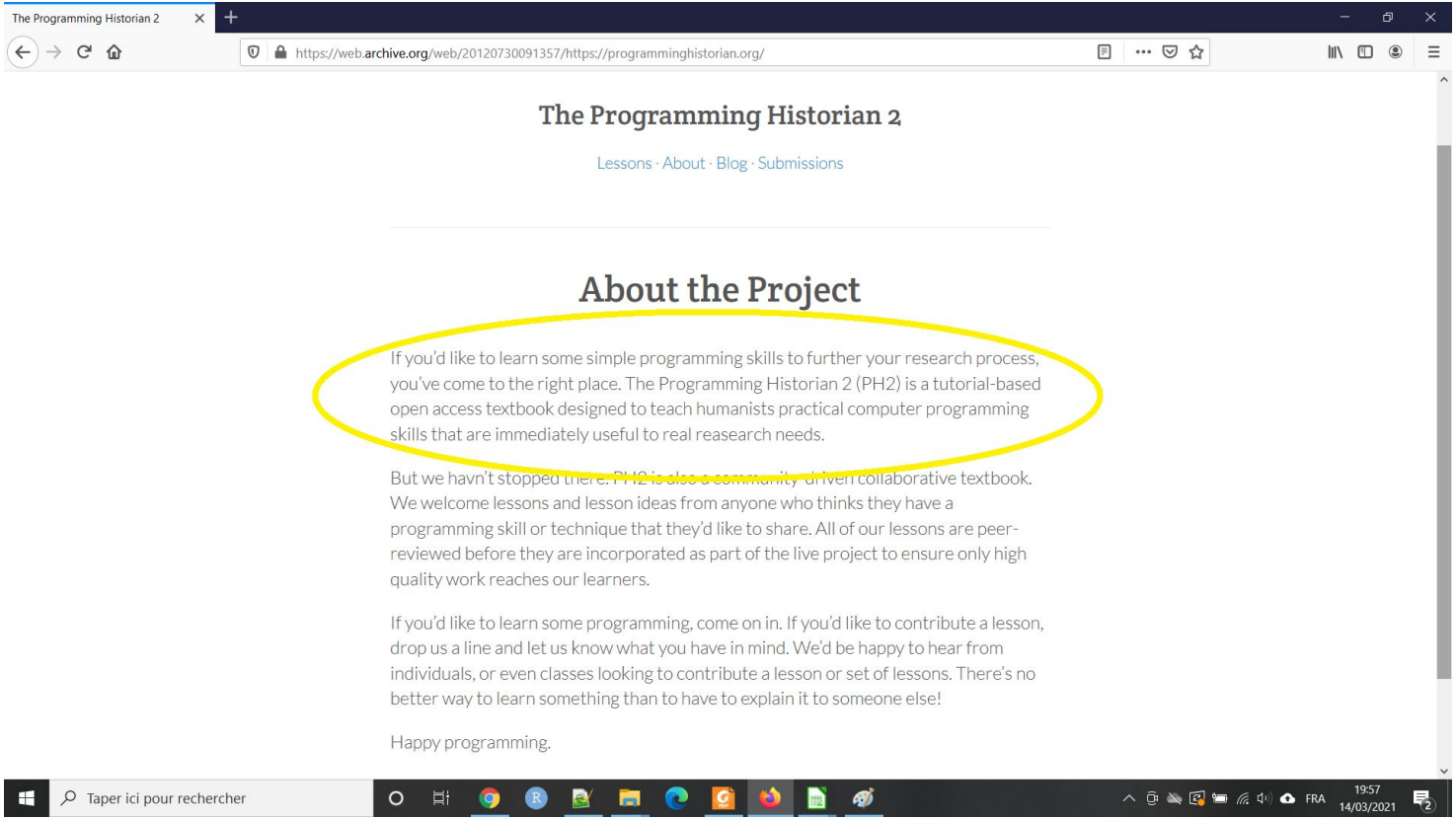
A Community-Based Approach

A Community of Practice (1)

A shared domain of interest

⇒ Building of programming skills, appropriation of digital methods/tools/workflows in research and teaching in history and the humanities

Part of the homepage of the PH website, 14 March 2021



Screenshot of the About page of the PH website in 2012 (Internet Archive)

A Community of Practice (2)

A community of engagement and sharing

- a core of volunteer editors who lead and run a scholarly journal
- a broader community of contributors who irrigate the project

Volunteer-driven project

- 28 volunteer editors by March 2021
- A total of 46 volunteer editors from 2008 to present

Source: Project Team page: https://programminghistorian.org/en/project-team

(Broader) Community-driven project

- Write a lesson
- Edit a lesson
- Translate a lesson
- Review a lesson
- Provide feedback/Report problems/Make suggestions
- Support Programming Historian
  - institutional partnerships
  - individual support (donations)
  - in kind services or project-specific funding from institutions

Our Supporters

The Programming Historian is grateful to our past and current supporters for enabling us to bring these resources to the world, free at the point of access to readers everywhere.

Institutional Partners

Contributers to our Institutional Partner Programme.

- KU Leuven Libraries, Belgium
- Institute of Historical Research Wohl Library, United Kingdom
- University of Sussex Library, United Kingdom
- Centre for Data, Culture and Society, University of Edinburgh, United Kingdom
- Western University Library, Canada
- UCL Centre for Digital Humanities, United Kingdom
- Corporation for Digital Scholarship, United States
- Université du Luxembourg, Luxembourg
- Software Sustainability Institute, United Kingdom

We welcome enquiries from prospective Institutional Partner Programme contributors.

Institutional Partnerships

Individual Supporters

Thank you for your interest supporting *Programming Historian*. Individual donors are vital for enabling our continued work, and for helping us to keep our content free to readers around the world. With one in three of our readers living in low and middle income countries, you're also helping to level the playing field, ensuring everyone is empowered to master technology and put it to good use, no matter where they live.

Your support directly enables the infrastructure that keeps our publications together. That means our volunteers can devote their time to maintaining our growing suite of lessons, and ensuring that new lessons are built to stand the test of time.

Ongoing Support

For $1, $5, or $15 per month, you can join our growing list of individual Patreon sponsors. By joining our Patreon supporters you enable our team to plan for the future and to grow our suite of digital humanities and digital skills publications.

Become a patron

Individual Supporters

Additional Supporters

Organizations that have or continue to give support ranging from in kind services, to project-specific funding.

- The University of Sussex, United Kingdom:
  - University of Sussex Library [2020-Present]
  - School of History, Art History and Philosophy, University of Sussex [2019]
  - ESRC Impact Accelerator Account, University of Sussex, [2019].
- The Roy Rosenzweig Center for New Media (RRCHNM), United States [2011-Present].
- The British Academy Writing Workshops funding scheme, United Kingdom [2018].
- The dSHARP lab at Carnegie Mellon University sponsored by the Andrew W. Mellon Foundation, United States [2018-2020].
- The Network in Canadian History & Environment, Canada [2011-2013].
- The Open Knowledge Foundation Open Data Day mini-grant scheme, United Kingdom [2021].

In kind support, project-specific funding

Everyone is credited for their contribution

Analyse de corpus avec AntConc

Heather Froehlich

L'analyse de corpus est un type d'analyse textuelle qui permet de faire des comparaisons d'objets textuels à grande échelle (la fameuse "lecture à distance" (*distant reading*)).

Évaluée par les pairs CC-BY 4.0 Soutenir PH

SUIVI ÉDITORIAL PAR
Fred Gibbs

ÉVALUATION PAR
Nabeel Siddiqui
Rob Sieczkiewicz

TRADUCTION PAR
Hugo Bonin

SUIVI DE TRADUCTION PAR
Sofia Papastamkou

ÉVALUATION DE TRADUCTION PAR
Antoine Champigny
François Dominic Laramée

Building Community

AUGUST 8, 2018

Writing Workshop Report Bogotá, Colombia

ADAM CRYMBLE AND MARIA JOSÉ AFANADOR-LLACH

Group Photograph of workshop participants in Bogota, 2018.

A Community of Practice (3)

A shared practice: resource sharing

“Peer-reviewed tutorials that help humanists learn a wide range of digital tools, techniques, and workflows”

Lessons: 6 skills

- Acquire (12)
- Transform (30)
- Analyze (17)
- Present (18)
- Sustain (3)

Lessons: 11 topics

APIS (6)

PYTHON (20)

DATA MANAGEMENT (8)

DATA MANIPULATION (23+1)

DISTANT READING (12+3)

SET UP (7)

LINKED OPEN DATA (1)

MAPPING (11+1)

NETWORK ANALYSIS (5)

WEB SCRAPING (6)

DIGITAL PUBLISHING (10)

HEATHER FROEHLICH

Corpus Analysis with Antconc

... ial: Working with Plain Text Files Getting Started w ..... Corpus analysis is a form of text analysis which allows you to ..... particular author or kind of text, particular kinds of grammatic ...

ZOË WILKINSON SALDAÑA

Sentiment Analysis for Exploratory Data Analysis

... nerate sentiment scores for a text. Critically evaluate the sen ..... ." (Tukey 1977:3)Exploring Text with Sentiment AnalysisWhen ..... categorizing and quantifying text. These NLP approaches, which i ...

SHAWN GRAHAM, SCOTT WEINGART, AND IAN MILLIGAN

Getting Started with Topic Modeling and MALLET

... it can be used on a corpus of texts to identify topicsfound in t ..... modeling tool takes a single text (or corpus) and looks forpat ..... rams that extracttopics from texts. A topic to the computer is a ...

FRANÇOIS DOMINIC LARAMÉE

Introduction to stylometry with Python

... others prefer long blocks of text consisting of many clauses. ..... writing style within a single text may indicate plagiarism,5 and ..... tribution. Given an anonymous text, it is sometimes possible to g ...

ANDREW AKHLAGHI

OCR and Machine Translation

... OCR and Machine TranslationCont ..... ick OCR Translation Putting it al ..... optical character recognition (OCR) and machine translation into ...

JON CRUMP

Generating an Ordered Data Set from an OCR Text File

... g an Ordered Data Set from an OCR Text FileGenerating an Order ..... tes strategies for taking raw OCR output from a scanned text, p ..... y discover that even the best OCR results in unacceptably high ...

LAURA TURNER O'HARA

Cleaning OCR'd text with Regular Expressions

... ut we've all noticed that the OCR for historictexts is far fro ..... s and formats make for unique OCR. Take for example, this page f ..... he PDF page.However, the OCR layer (downloaded as a text f ...

MORITZ MÄHR

Working with batches of PDF files

... Optical Character Recognition (OCR) Extract embedded texts from ..... pics it containsTesseract OCR software used in this lesson ..... pt install ocrmypdf tesseract- ocr -all poppler-utils imagemagick ...

Análisis de corpus con Voyant Tools

Silvia Gutiérrez De la Torre

En este tutorial se aprenderá cómo organizar y analizar un conjunto de textos con Voyant-Tools.

Revisado por pares CC-BY 4.0 Apoyar PH

EDITADO POR
Jennifer Isasi

REVISADO POR
Daniela Ávido
Jennifer Isasi

PUBLICADO 2019-04-20 MODIFICADO 2020-05-12 DIFICULTAD Bajo https://doi.org/10.46430/phes0043

Lessons: Uses/Impact

- Self-training
- University syllabi
- Libraries and archives (cf. Library Carpentries)
- Beyond humanities

Elias W. BA
@eliaswalyba

Est-ce que vous connaissez le site Programming Historian (programminghistorian.org) ? C'est une excellente ressource pour apprendre à programmer. Les cours sont gratuites et disponibles en 3 langues: anglais, espagnol et français. Testez pour voir.

Translate Tweet

2:38 AM · May 2, 2020 · Twitter Web App

37 Retweets **1** Quote Tweet **70** Likes

Relevant people

Elias W. BA
@eliaswalyba

Following

Proud Dad • Social Entrepreneur • Programmer • Lead Data Scientist • Machine Learning Teacher • Co-Founder & Lead @GalsenAI & @NeographSchool

Paris trends

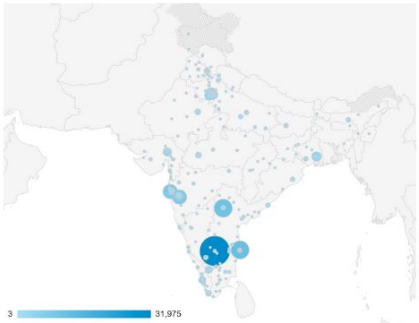
#Cherry
Exclusivement sur Apple TV+

Tweet of @eliaswalyba, dev, Dakar, Senegal

APRIL 4, 2018

The Programming Historian and India

ADAM CRYMBLE AND JAMES BAKER



Traffic Report of Programming Historian visitors from India, 2012-2018.

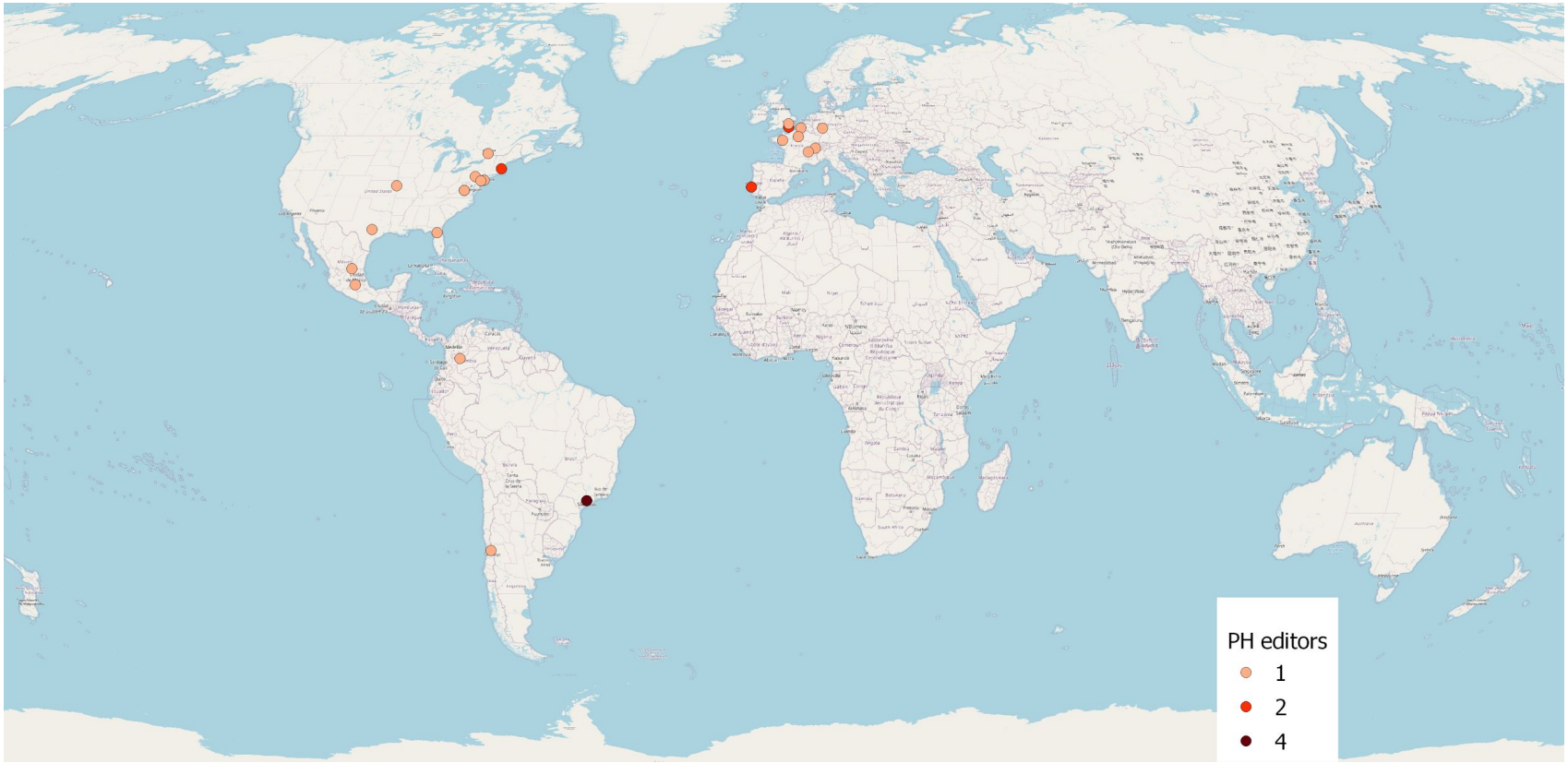
When the *Programming Historian* launched its open access tutorials in 2012, historians were the target audience. By 2014 our audience statistics had already presented a surprise: India had emerged as the second largest source of *Programming Historian* readers - a title it still holds in 2018.

Country	Visitors 2017	Per cent Increase from 2016
United States	100,853	219
India	43,906	276
Great Britain	24,688	252
Canada	11,575	215
Germany	11,645	225

While India is home to many talented historians, we suspect the project may have attracted a very welcome but unintended audience. The traffic is concentrated in four cities, topped by Bengaluru, India's equivalent of Silicon Valley. In fact, four of the top ten cities in the world for *Programming Historian* traffic, are all in India:

Top Ten Cities	Traffic Volume 2017
Bengaluru, India	18,222
London, UK	13,113
New York, USA	12,102

Adam Crymble, James Baker, [“The PH and India”](), April 4, 2018



PH editors on March 16, 2021

Multilingualism

[PH]

Enter *The Programming Historian*

(The initial English version)

84 lessons ISSN: 2397-2068

We publish novice-friendly, peer-reviewed tutorials that help humanists learn a wide range of digital tools, techniques, and workflows to facilitate research and teaching.

[P̃H]

Entrar *The Programming Historian en español*

48 lecciones ISSN: 2517-5769

Publicamos tutoriales revisados por pares dirigidos a humanistas que quieran aprender una amplia gama de herramientas digitales, técnicas computacionales y flujos de trabajo útiles para investigar y enseñar.

[P̂H]

Entrez *The Programming Historian en français*

15 leçons ISSN: 2631-9462

Nous publions des tutoriels évalués par des pairs qui permettent l'initiation à et l'apprentissage d'un large éventail d'outils numériques, de techniques et de flux de travail pour faciliter la recherche et l'enseignement en sciences humaines et sociales.

[PtH]

Entrar *The Programming Historian em português*

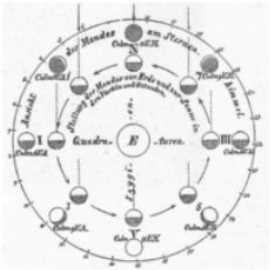
1 lições ISSN: 2753-9296

Publicamos tutoriais acessíveis, avaliados por pares, que ajudam os humanistas a aprender uma ampla gama de ferramentas digitais, técnicas computacionais e metodologias de trabalho que facilitam a pesquisa e o ensino.

[PH]

The Programming Historian
日本語訳

このページでは、欧米のデジタルヒストリー研究ノウハウ公開・共有プロジェクトであるThe Programming Historianのリソースの一部を、日本語に翻訳公開しています。今後も、順次コンテンツを増やしていきます。



2020年5月9日 2020年6月24日
Palladio, ネットワーク分析, 中級, 可視化

解釈学からデータのネットワークへ：史料からのデータ抽出とネットワークの可視化

Marten Düring ネットワークの可視化は、人文系の研究者がテキスト史料

続きを読む ›

人気の記事

Wgetを使った自動ダウンロード法

解釈学からデータのネットワークへ：史料か...

第2章 TEIヘッダー

Preservar os seus dados de investigação

James Baker

Esta lição irá sugerir maneiras pelas quais os historiadores podem documentar e estruturar os seus dados de pesquisa, a fim de garantir que continuem sendo acessíveis no futuro.

Avaliada por pares CC-BY 4.0 Apoie o PH

EDITADO POR Adam Crymble
REVISTO POR Jane Winters, Sharon Howard, William J. Turkel
TRADUZIDO POR Márcia T. Cavalcanti
TRADUÇÃO EDITADA POR Joana Vieira Paulino
TRADUÇÃO REVISTA POR Maria Guedes, Elizabeth Freire

PUBLICADO 2014-04-30 TRADUZIDO 2021-01-29 MODIFICADO 2021-01-29 DIFICULDADE Baixo https://doi.org/10.46430/phpt0001

Disponível em: EN (original) | ES | PT | FR

Conteúdos

Histórico

No ensaio de 2003 "Escassez ou abundância", Roy Rosenzweig procurou alertar os historiadores para o que ele chamou de "fragilidade das evidências na era digital" (Rosenzweig, 736). E, embora as suas preocupações estivessem focadas em fontes disponíveis na Web aberta, podem ser facilmente estendidas aos materiais digitais - ou dados - que os historiadores criam durante as suas pesquisas.

É nesses dados de pesquisa que o presente guia se focará. Mas por quê?

Author Guidelines

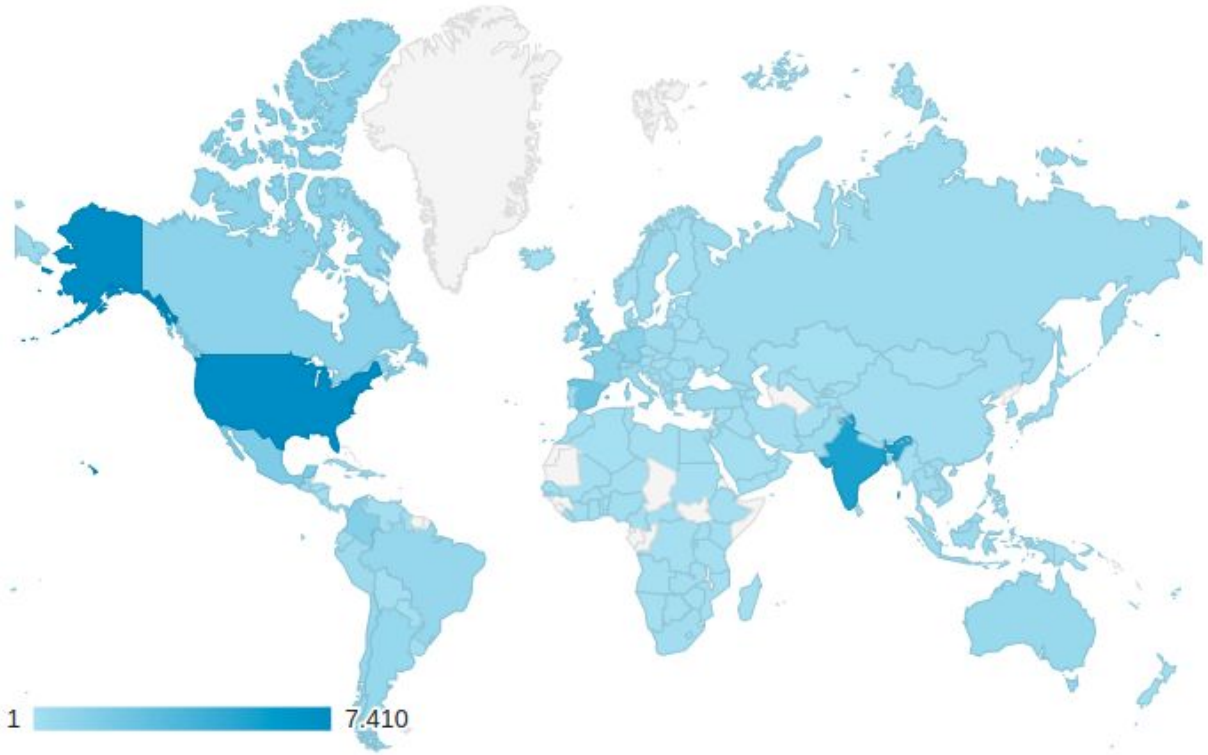
Write for a Global Audience

Programming Historian readers live all around the world. Authors can and should take steps to write their lesson accessibly for as many people as possible. Follow these global-facing guidelines:

Author Guidelines

- **Multi-lingual**: when choosing methods or tools, make choices with multi-lingual readers in mind – especially for textual analysis methods, which may not support other character sets or may only provide intellectually robust results when used on English texts. Where possible, choose approaches that have multi-lingual documentation, or provide multi-lingual references for further reading. This will help our translators.

Readers (last week)



Challenges of having different linguistic versions

- development and maintenance of flexible and evolutive infrastructure
- develop original lessons in all languages of *Programming Historian* to fulfill specific needs of the linguistic communities
- encourage translation **to** and not only **from** English
- encourage translation between the non-English linguistic versions of the PH
- recruit editors with linguistic skills and multicultural acquaintance to assure infra-team and trans-subteams effective collaboration

Our biggest challenge

Diversity!

- work with under-represented communities in the global DH
- further contribute to digital and code literacy through enrichment with needs and approaches of culturally diverse environments.
- address the English-centeredness of technologies.

Thank you!
¡Gracias!
Merci!
Obrigada!

And if you want to join the community, you are welcome!

https://programminghistorian.org/
@ProgHist