

# Internationalisation, localisation and translation crowd sourcing at Wikimedia Foundation

Siebrand Mazeland

Product Manager Language Engineering, Wikimedia Foundation

Community Manager, [translatewiki.net](http://translatewiki.net)



# Wiki what?

- ✦ Wikimedia Foundation
- ✦ Wikimedia Movement
- ✦ Wikipedia
- ✦ MediaWiki
- ✦ [translatewiki.net](http://translatewiki.net)

# Wikimedia Foundation



*Imagine a world in which every single human being can freely share in the sum of all knowledge. That's our commitment.*



# Wikimedia Foundation



- ✦ **Wikimedia Foundation** is the American non-profit that operates several online collaborative wiki projects including Wikipedia
- ✦ Created 2003
- ✦ Currently 135 employees
- ✦ USD 46M budget for 2012-2013 fiscal year
- ✦ Does not create content for the wiki projects



# Wikimedia movement

- ✦ The **Wikimedia movement** is the totality of people, organisations, activities and values which revolve around Wikimedia projects, or anyone who makes a valuable contribution to Wikimedia's free knowledge initiative
- ✦ Emerged around 2004 as Wikipedia and sister projects became more popular
- ✦ Has around 100,000 active volunteers every month

# Wikipedia



- ✦ **Wikipedia** is a project dedicated to the building of free encyclopedias in all languages of the world.
- ✦ Started with English language Wikipedia on January 15, 2001
- ✦ Available in 285 language versions
- ✦ 24.6M content pages
- ✦ ~500M unique monthly visitors
- ✦ ~22B monthly page views



# Hundreds is massive



Na Vosa Vakaviti English 𑌵𑌶𑌰𑌫𑌰 Araona Basa Banyumasan 𐤆𐤇𐤅 Vardá монгол ᠮᠣᠩᠭᠣᠯ ᠪᠣᠮ᠋ᠠᠨᠢ Romani meänkieli Yup'ik Gegë 𐌆𐌆𐌆𐌆𐌆𐌆 Iboran keel 𐌆𐌆𐌆𐌆𐌆𐌆  
norsk bokmål Chavacano de Zamboanga 𑌵𑌶𑌰𑌫𑌰 Malti 日本語 Sranantongo Silozi 𑌵𑌶𑌰𑌫𑌰 interlingua Bahasa Melayu corsu italiano Lojban  
𐌆𐌆𐌆𐌆𐌆𐌆 Test (site admin only) 𑌵𑌶𑌰𑌫𑌰 Kongo 𑌵𑌶𑌰𑌫𑌰 Emiliàn Ἀρχαία ἑλληνική Ελληνικά kaszëbsczi Gaelg Kinyarwanda Gagana Samoa Lëtzebuergesch  
𑌵𑌶𑌰𑌫𑌰 Limburgs Turoyo Ирон हिन्दी Võro Deutsch feroyskt татарча 𑌵𑌶𑌰𑌫𑌰 magyar tojikī 𑌵𑌶𑌰𑌫𑌰 العربية Bidama  
Bân-lâm-gú эрзянь 𑌵𑌶𑌰𑌫𑌰 asturianu tatarça Ripoarisch Sesotho Basa Jawa 𑌵𑌶𑌰𑌫𑌰 Bahasa Betawi Tagalog Volapük português  
Xitsonga Dorerin Naoero 𑌵𑌶𑌰𑌫𑌰 贛語 (繁體) euskara bamanankan azərbaycanca Tašlhiyt/𑌵𑌶𑌰𑌫𑌰 Alemannisch 𑌵𑌶𑌰𑌫𑌰 ئۇيغۇرچە lingála қазақша 𑌵𑌶𑌰𑌫𑌰  
עברית Plattdüütsch 中文 (香港) 𑌵𑌶𑌰𑌫𑌰 къарачай-малкъар Cymraeg galego žemaitėška Hausa Aymar aru Tok Pisin Kotava 𑌵𑌶𑌰𑌫𑌰  
српски (ћирилица) Igbo norsk bokmål Basa Sunda саха тыла 𑌵𑌶𑌰𑌫𑌰 Qaraqalpaqsha Rumagnòl Akan isiZulu hrvatski Faaag Rotuma  
नेपाली slovenščina Ilokano 𑌵𑌶𑌰𑌫𑌰 português do Brasil norsk nynorsk aragonés Lingua Franca Nova 𑌵𑌶𑌰𑌫𑌰 한국어 (조선) Bahasa Banjar  
Interlingue Frysk Mağribi dansk 𑌵𑌶𑌰𑌫𑌰 dolnoserbski українська Nederlands Santali ckb-arab occitan Diné bizaad Àarjelsaemien  
𑌵𑌶𑌰𑌫𑌰 Winaray Konknni 𑌵𑌶𑌰𑌫𑌰 اردو Schläsch 𑌵𑌶𑌰𑌫𑌰 Cmique Itom rumantsch Albaamo innaaililka Gagauz British English Magyar (magázó) Scots emiliàn e rumagnòl  
𑌵𑌶𑌰𑌫𑌰 中文 (新加坡) Pälzisch Boarisch 𑌵𑌶𑌰𑌫𑌰 بختیاری 𑌵𑌶𑌰𑌫𑌰 Österreichisches Deutsch 𑌵𑌶𑌰𑌫𑌰 فارسی Toki Pona Fulfulde Nedersaksies русиньскый 𑌵𑌶𑌰𑌫𑌰 中文 (中国大陆)  
𑌵𑌶𑌰𑌫𑌰 𑌵𑌶𑌰𑌫𑌰 𑌵𑌶𑌰𑌫𑌰 Karufi Säggesch 𑌵𑌶𑌰𑌫𑌰 қазық (Қазақстан) Bafa walon žemaitėška Cebuano furlan Адыгэбзэ tetun точыкй Kurdî Mi'kmaq Plautdietsch Sassaresu Livõ keļ  
qirimtatarca (Latin) мокшень mapudungun Wayúu 𑌵𑌶𑌰𑌫𑌰 𑌵𑌶𑌰𑌫𑌰 Maayr Taan беларуская (тарашкевіца) 中文 𑌵𑌶𑌰𑌫𑌰 অসমীয়া 𑌵𑌶𑌰𑌫𑌰 беларуская (тарашкевіца)  
lea faka-Tonga Bahasa Selayar Kirmancki 𑌵𑌶𑌰𑌫𑌰 català Kreyòl ayisyen sicilianu Mìng-dĕng-ngŭj Kinárny-a Betak Toba 𑌵𑌶𑌰𑌫𑌰 文言 Влакесте Kurdî (latîni) буряад Ladino  
Luganda Prūsiskan Hawai'i Vahcuengh Gaeilge македонски 𑌵𑌶𑌰𑌫𑌰 湘语 Runa shimi Seeltersk eesti Malagasy 𑌵𑌶𑌰𑌫𑌰 Uyghurche 𑌵𑌶𑌰𑌫𑌰 Аньсаша Kapampangan  
Gàidhlig 𑌵𑌶𑌰𑌫𑌰 Кыргызча толышə зывон vepsän kel' lumbaart олык марий kalaallisut latgalu Tiếng Việt český Mirandés  
Pangasinan ПаллгӀай Musi Runa Simi 𑌵𑌶𑌰𑌫𑌰 Nouormand tarandíne хальмг кыырымтатарджа (Кирилл) Picard 𑌵𑌶𑌰𑌫𑌰 ku-laku-laku-laku-laku 𑌵𑌶𑌰𑌫𑌰 বিষ্ণুপ্রিয়া মণিপুরী भोजपुरी  
удмурт arpetan Hak-kâ-fa 中文 (繁體) română Пере́м Коми Cariceño srpski (latinica) 𑌵𑌶𑌰𑌫𑌰 слова́ньскъ / 𑌵𑌶𑌰𑌫𑌰  
español de America Latina nl-be संस्कृतम् Avañe'ê 𑌵𑌶𑌰𑌫𑌰 o'zbekcha 𑌵𑌶𑌰𑌫𑌰 Kiswahili молдовеняскэ egegbe français 𑌵𑌶𑌰𑌫𑌰 español  
Türkmençe 𑌵𑌶𑌰𑌫𑌰 বাংলা مصرى Afrikaans qazaqşa (latin) Piemontèis қазақша (кирил) slovenčina Uyghurche  
srpskohrvatski / српскохрватски 𑌵𑌶𑌰𑌫𑌰 ko e vagahau Niue 𑌵𑌶𑌰𑌫𑌰 Setswana latviešu 𑌵𑌶𑌰𑌫𑌰 粵語 Lazurî polski Soomaaliga Novial Чӕвашла 𑌵𑌶𑌰𑌫𑌰 འདྲེན་ལོག་ Wolof 𑌵𑌶𑌰𑌫𑌰 русский  
Canadian English 𑌵𑌶𑌰𑌫𑌰 قازاقشا 中文 (台灣) SiSwati 𑌵𑌶𑌰𑌫𑌰 ਪੰਜਾਬੀ Esperanto Bráhuī 𑌵𑌶𑌰𑌫𑌰 Armãeashce Patois 𑌵𑌶𑌰𑌫𑌰 贛語 (简体) лезги Acèh  
Nederlands (informeel) Sängö нохчийн Tosk Message documentation Nahuatl 𑌵𑌶𑌰𑌫𑌰 नेपाल भाषा American sign language 𑌵𑌶𑌰𑌫𑌰 Māori кырык мары Dziri  
Адыгэбзэ Deutsch West-Vlams भोजपुरी Nnapulitano српски / srpski Baso Minangkabau Zazaki Sesotho sa Leboa inuktitut Nordfriisk lietuvij  
Bikol Central коми 𑌵𑌶𑌰𑌫𑌰 𑌵𑌶𑌰𑌫𑌰 suomi Krio kernowek 𑌵𑌶𑌰𑌫𑌰 Norfolk / Pitkern татарча/татарça 𑌵𑌶𑌰𑌫𑌰 Yorùbá Deutsch (Sie-Form) arero rapa nui 𑌵𑌶𑌰𑌫𑌰 ไทย  
башҡортса 𑌵𑌶𑌰𑌫𑌰 𑌵𑌶𑌰𑌫𑌰 پښتو کوردی sámegiella vèneto Fiji Hindi تورکجه Chamoru Ido 𑌵𑌶𑌰𑌫𑌰 ku-laku-latin Dusun Bundu-liwan 𑌵𑌶𑌰𑌫𑌰 chiShona brezhoneg Gothic  
Papiamentu pt-pt Tarifit Ænglisc Mainfränkisch 𑌵𑌶𑌰𑌫𑌰 한국어 Ποντιακά isiXhosa Bahasa Indonesia Ligure Zeëuws Karjala bosanski 𑌵𑌶𑌰𑌫𑌰 𑌵𑌶𑌰𑌫𑌰 بلوچی مکرانی 𑌵𑌶𑌰𑌫𑌰  
šlůnski 𑌵𑌶𑌰𑌫𑌰 sán-lám-gú estremeñu viãhepe sardu llonggo тыва дыл मैथिली 𑌵𑌶𑌰𑌫𑌰 مازرونی 中文 (简体) मराठी íslenska 𑌵𑌶𑌰𑌫𑌰 𑌵𑌶𑌰𑌫𑌰 hornjoserbsce  
Mizo ꠘawng 𑌵𑌶𑌰𑌫𑌰 پنجابی Latina Guadeloupean Creole French français cadien zza беларуская Schweizer Hochdeutsch български Türkçe 𑌵𑌶𑌰𑌫𑌰 shqip  
Taqbaylit español (formal) авар svenska



# MediaWiki



- ✦ **MediaWiki** is a free wiki software application, developed by the Wikimedia Foundation and others
- ✦ Used to run all of the projects hosted by the Foundation, including Wikipedia
- ✦ Highly customisable, with more than 700 configuration settings and more than 1,800 extensions
- ✦ User interface has been fully or partially translated into more than 300 languages



# translatewiki.net



- ✦ **translatewiki.net** the localisation platform for translation communities, language communities, and free and open source projects
- ✦ Supports online and offline translation for MediaWiki and other software
- ✦ Started by Niklas Laxström in 2005
- ✦ 4,700 registered translators, 500 active/month
- ✦ 60,000 strings translated/month



# Internationalisation

- ✦ Allow localisation of our sites in every aspect
- ✦ Formerly done mostly by volunteers, still very important
- ✦ Wikimedia Language Engineering team formed in 2011



વિકિપીડિયા  
મુક્ત વિશ્વજ્ઞાનકોષ

# Messaging framework

## Internationalisation

- ✦ Make as flexible a string localisation framework without making it really hard on translators
- ✦ Works in PHP and JS
- ✦ Plural, gender, grammar
- ✦ Formal/informal planned

- ✦ Date/time, numbers, URLs, ...

String example:

```
{{GENDER:$1|He|She|They}} bought {{PLURAL:$2|0=no boat|one boat|$2 boats|12=a dozen boats}} and got {{PLURAL:$3|0=no|only one|$3 USD}} discount.
```

# Internationalisation

- ✦ Generic JavaScript libraries for common i18n challenges
  - ✦ Messaging
  - ✦ Language selection
  - ✦ Input methods
  - ✦ Web fonts

All code is on GitHub!

- ✦ [Wikimedia/jquery.i18n](#)
- ✦ [Wikimedia/jquery.uls](#)
- ✦ [Wikimedia/jquery.ime](#)
- ✦ [Wikimedia/jquery.webfonts](#)

Use, fork, improve!

# Language Selector

## Internationalisation

- ✦ How to choose from 400 languages?
  - ✦ Smart display with rows and columns
  - ✦ Use accept language
  - ✦ Geo location
  - ✦ Grouping by region
- ✦ Grouping by script
- ✦ Typos allows in search
- ✦ Language code in search
- ✦ Auto completion
- ✦ Cross language search with CLDR data

# Language Selector

Select Language ✕

Worldwide America Europe Middle East Africa Asia Pacific


Language search

Common languages

Deutsch	français	Nederlands	GWY
Diné bizaad	Hawai'i	Tiếng Việt	中文 (繁體)
English	italiano	русский	한국어
español	Iñupiak	עברית	

Worldwide

English	Lingua Franca Nova
español	Lojban
Esperanto	Novial
français	português
Ido	Toki Pona

Aa Display settings  Input settings

# Input methods

## Internationalisation

- ✦ Allow input of any language and any script with a US keyboard
  - ✦ 155+ input methods
  - ✦ 75+ languages
  - ✦ Largest open source collection usable on any website



# Web fonts

## Internationalisation

- ✦ No more tofu blocks on web pages
  - ✦ JavaScript and CSS3 cause download of font based on language tag in content
  - ✦ Supports any font per configuration
- ✦ Wikimedia only uses open licensed fonts
  - ✦ 44 fonts
  - ✦ 20 scripts



# jQuery.i18n

## Internationalisation

- ✦ jQuery i18n framework
  - ✦ JSON file format
  - ✦ ECMA Standard
  - ✦ Message Parser w/  
PLURAL & GENDER
  - ✦ Grammar support
- ✦ To come
  - ✦ Date/time formatting
  - ✦ Number formatting

# Localisation

## translatewiki.net

- ✦ Central place to translate MediaWiki and extensions
- ✦ Localise wiki style
- ✦ Assume good faith
- ✦ Translator has central role
- ✦ Instant deployment (live next day on Wikipedia)



# Provide adequate tooling

- ✦ MediaWiki extension Translate
- ✦ Translation Memory based on Solr
- ✦ Machine translation helpers (Google, Microsoft, Yandex, Apertium)
- ✦ User generated message documentation
- ✦ Editor recently redesigned
- ✦ Respond quickly to translator queries

# Localisation

Message group [All](#) ▸ [Article Feedback v5](#)

Translate to [Nederlands](#) ▾

Article feedback

[All](#) [Untranslated](#) [Outdated](#) [Translated](#) ...



The submitted feedback is a duplicate of the last feedback; perhaps you accidentally submitted th...

[/ Edit](#)

Useful [Uitgelicht](#)

Outdated

[/ Edit](#)

No action needed

[/ Edit](#)

Inappropriate [Verborgen](#)

Outdated

[/ Edit](#)

Archived

[/ Edit](#)

Mark as:

[/ Edit](#)



# Localisation

Ordered list Nummerierte Liste ✓ Proofread / Edit

Unordered List Ungeordnete Liste ✓ Proofread / Edit

EtherpadLite:Pad.toolbar.indent.title/de ↗ ✕

Indent

Einrücken

**Save translation** **Skip to next**

Press "CTRL-OPTION-S" to save or "CTRL-OPTION-D" to skip to next message

**Used as tooltip of button**

**Suggestions**

Einrücken	100% match
Gedankenstrich <a href="#">Use as translation</a>	Yandex

**In other languages**

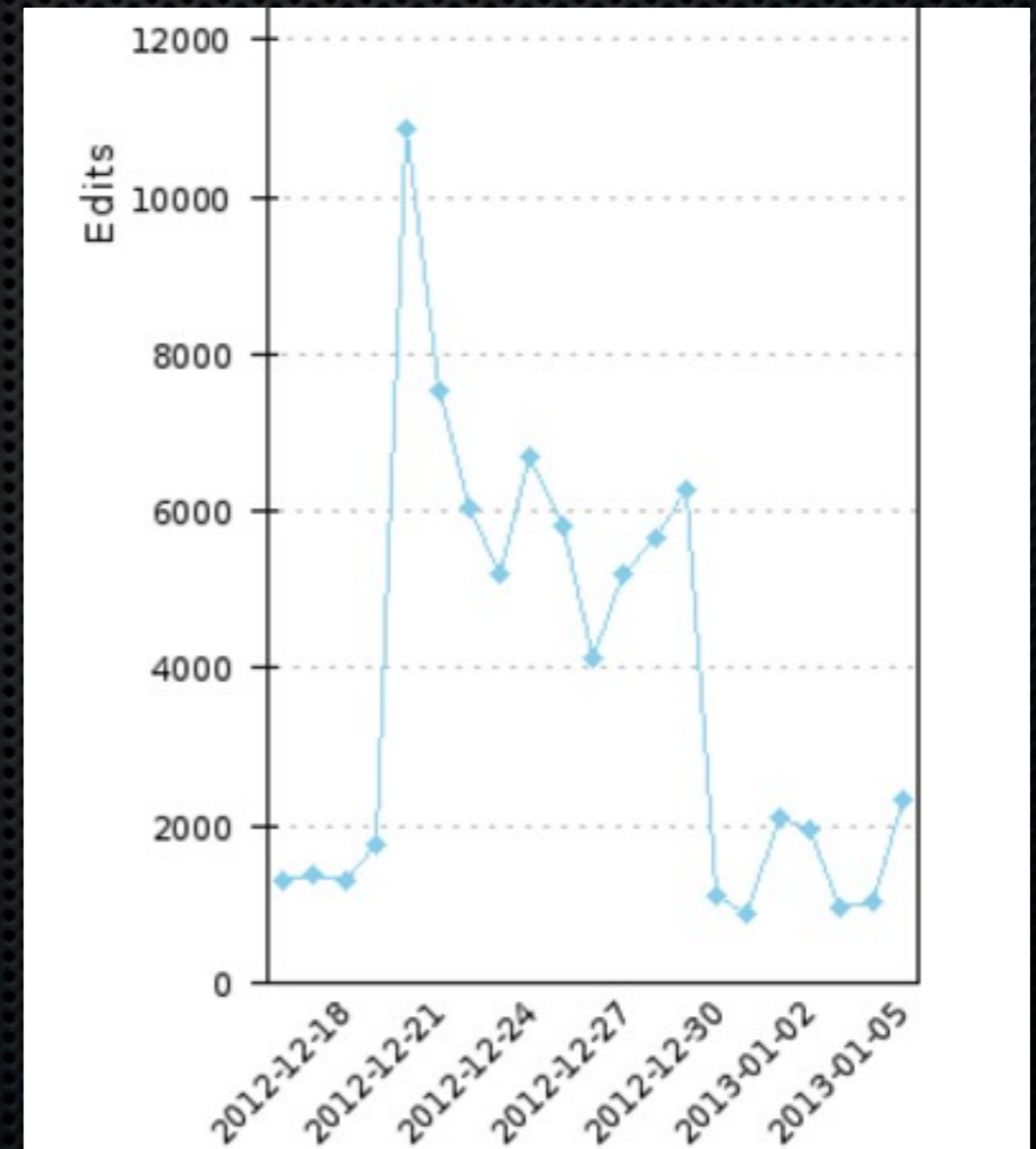
Inspringen	Nederlands
------------	------------

Outdent Ausrücken ✓ Proofread / Edit

Undo (Ctrl-Z) Rückgängig (Strg-Z) ✓ Proofread / Edit

# Localisation translatewiki.net

- ✦ Incentivise users
  - ✦ See result of a 1,000 Euro pledge
- ✦ Regular newsletters encourage regular activity
- ✦ Need more sponsors ;)



# Translation

- ✦ Policy
- ✦ Documentation
- ✦ Announcements
- ✦ Reports
- ✦ Proposals
- ✦ Translators sign up
- ✦ Targeted notifications

# Translation

## Translator signup

Use this page to indicate what languages you can translate in, and how you want to be contacted about new translation requests.

### User information

Username: Siebrand  
Email status: Your email address is confirmed

### Languages

Language #1

Language #2

Language #3

### Preferred contact methods

- Email
- Talk page
- Talk page on other wiki

### Contact frequency

- When there is something new to translate
- At most once a week
- At most once a month
- Weekly digest
- Monthly digest



# Translation

## Vision

[Translate this page](#)

### Other languages:

[Arabic](#) • [Asturian](#) • [Bulgarian](#) • [Catalan](#) • [Czech](#) • [German](#) • [Greek](#) • **[English](#)** • [Spanish](#) • [Extremaduran](#) • [Finnish](#) • [French](#) • [Galician](#) • [Swiss German](#) • [Hebrew](#) • [Croatian](#) • [Hungarian](#) • [Interlingua](#) • [Indonesian](#) • [Italian](#) • [Japanese](#) • [Korean](#) • [Lithuanian](#) • [Macedonian](#) • [Malayalam](#) • [Marathi](#) • [Norwegian Bokmål](#) • [Dutch](#) • [Polish](#) • [Portuguese](#) • [Brazilian Portuguese](#) • [Russian](#) • [Thai](#) • [Tagalog](#) • [Ukrainian](#) • [Vietnamese](#)

The **vision statement** of the Wikimedia Foundation describes our dreams, hopes and ambitions; our most radical conception of our organization and community — 20, 50, 100 years from today. It stands in contrast to the *mission statement*, which aims to be a more realistic description of the status quo. Our current vision statement is:

Imagine a world in which every single human being can freely share in the sum of all knowledge. That's our commitment.

# Translation

## Идеалы

This page is a [translated version](#) of a page [Vision](#) and the translation is 83% complete.

### Other languages:

[Arabic](#) • [Asturian](#) • [Bulgarian](#) • [Catalan](#) • [Czech](#) • [German](#) • [Greek](#) • [English](#) • [Spanish](#) • [Extremaduran](#) • [Finnish](#) • [French](#) • [Galician](#) • [Swiss German](#) • [Hebrew](#) • [Croatian](#) • [Hungarian](#) • [Interlingua](#) • [Indonesian](#) • [Italian](#) • [Japanese](#) • [Korean](#) • [Lithuanian](#) • [Macedonian](#) • [Malayalam](#) • [Marathi](#) • [Norwegian Bokmål](#) • [Dutch](#) • [Polish](#) • [Portuguese](#) • [Brazilian Portuguese](#) • [Russian](#) • [Thai](#) • [Tagalog](#) • [Ukrainian](#) • [Vietnamese](#)

**Формулировка идеалов** Фонда Викимедиа описывает наши мечты, надежды и стремления; основные представления о нашей организации и сообществе на 20, 50, 100 лет вперёд. Идеалы несколько отличаются от *определения миссии*, ориентированного на более реалистичное описание текущего состояния. Наше текущее представление об идеалах формулируется следующим образом:

Представьте себе мир, в котором каждый человек обладает свободным доступом ко всем накопленным человечеством знаниям. Мы создаём такой мир.

Предложения по изменению этой формулировки следует оставлять на странице [Vision/Unstable](#), все предложения будут рассматриваться, решения будут приниматься по крайней мере ежегодно.

# That's the introduction

- ✦ [siebrand@wikimedia.org](mailto:siebrand@wikimedia.org)
- ✦ [irc: #mediawiki-i18n at freenode.net](irc://freenode.net/#mediawiki-i18n)
- ✦ [hiring: wikimediafoundation.org/wiki/Jobs](http://wikimediafoundation.org/wiki/Jobs)
  
- ✦ Q&A, discussion