Microsoft® Research
Faculty Summit 2010

Planet-Wiki? Translation, Language and Power on Wikipedia

Amit Ray, Ph.D.
Associate Professor, Department of English
Rochester Institute of Technology

PART ONE

The Humanities and Digitality

Humanity

Species

Language

What characterizes a Language?

- grammar: the structural relationships between the signs in a system
- pragmatics: the context within which signs are used
- semantics: the content and context which give signs meaning

Orality/Presence

Sumerian Tablet
circa 2400 BCE
Cary Collection
RIT

Writing/
Re-presentation

Humanities

History
Literature
Philosophy

Reading and Writing

Encoding and Decoding Signs

The Linguistic Sign functions through Analogy

Analog Code

When the epoch of analog (which was to say also of the richness of language, of analogy) was giving way to the digital era, the final victory of the numerate over the literate."

Salman Rushdie, *Fury* (2001)

Analog and Binary Code

2010

circa 2400 BCE

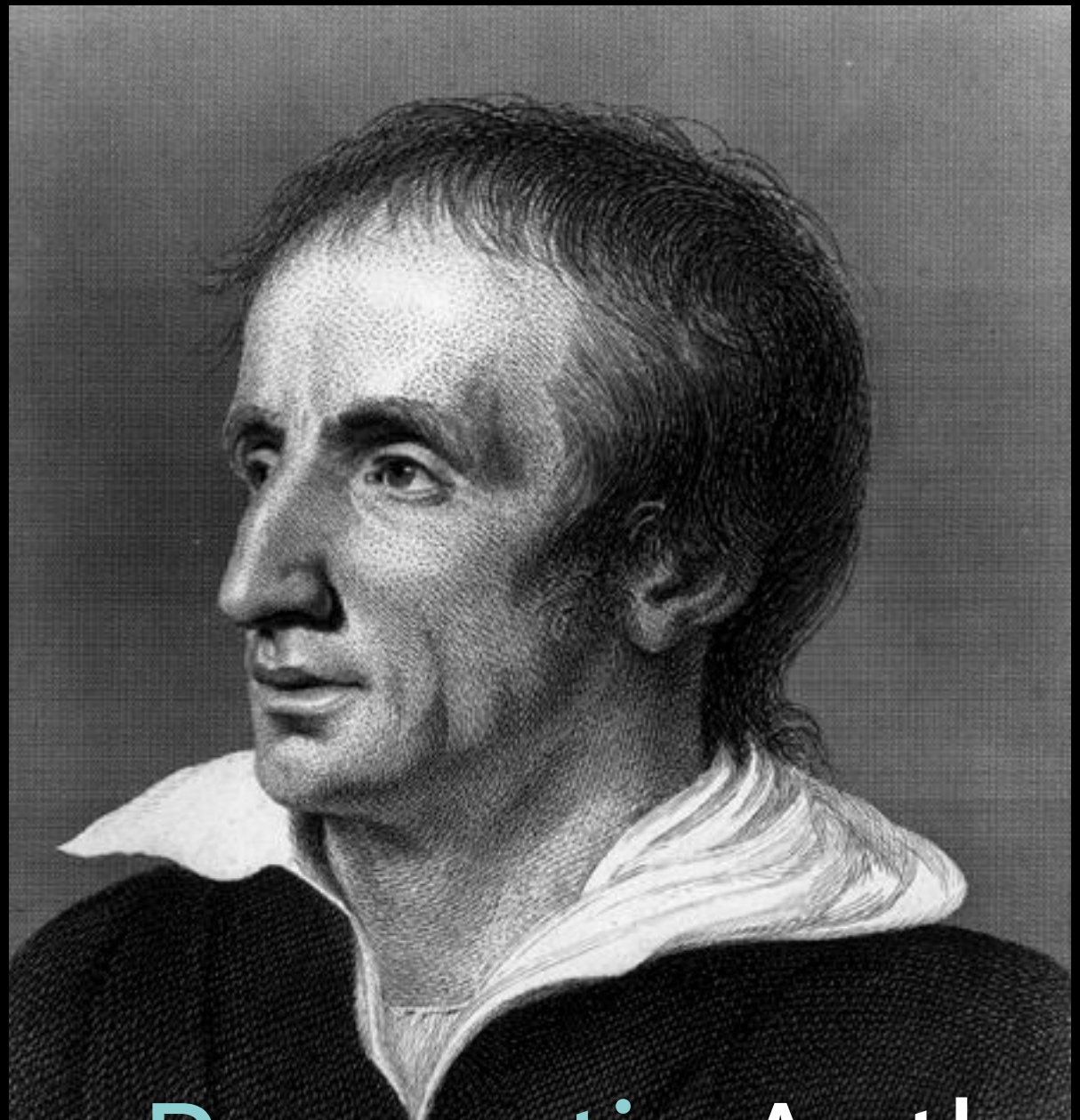
Cary Collection
RIT

PART TWO

Wikipedia and Distributed Authorship

What is an Author ?

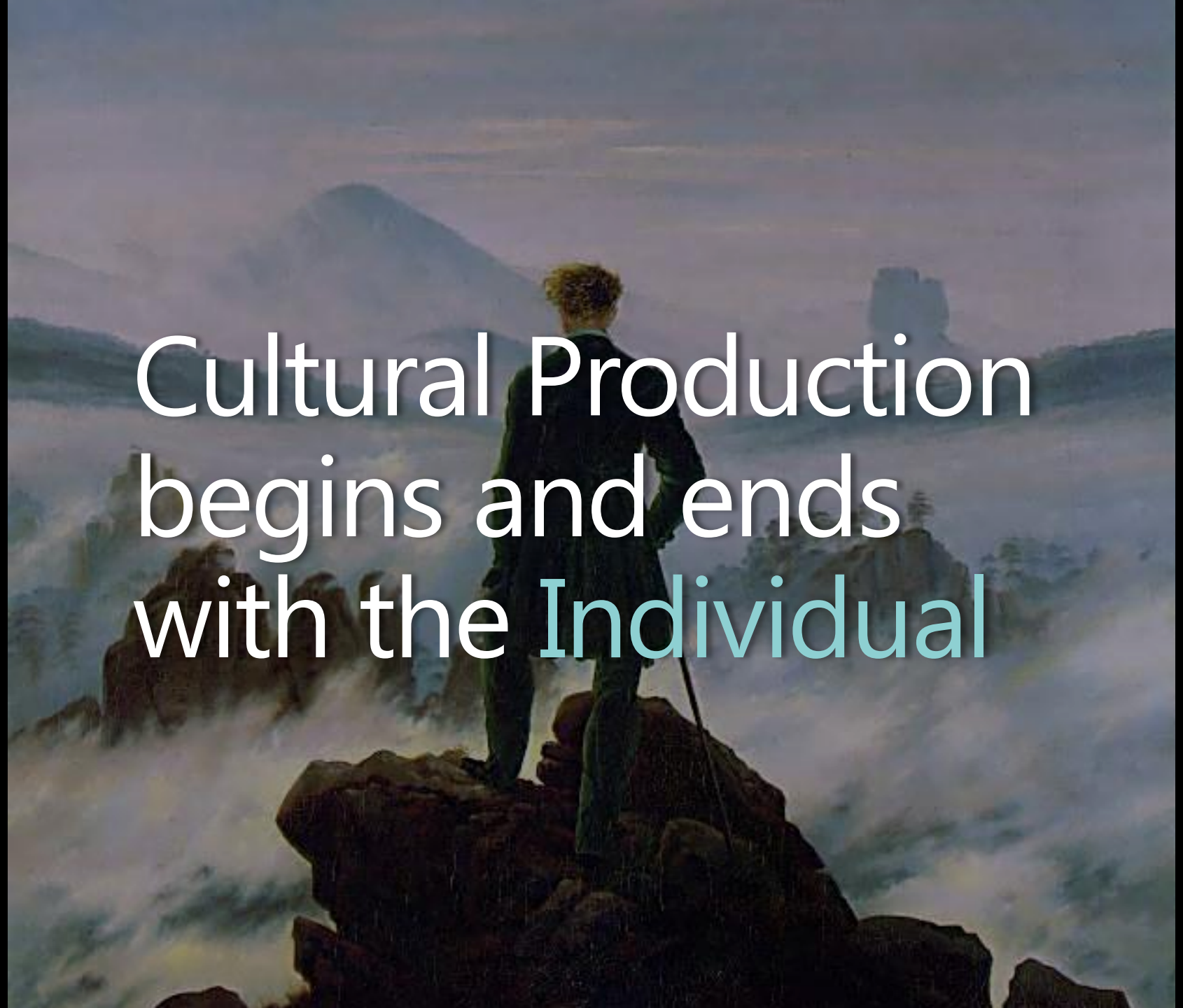
Authorship and Identity



The Romantic Author

Identity and Creation

creation
authorship
individuality
property


Cultural Production
begins and ends
with the Individual

Culture of the Individual

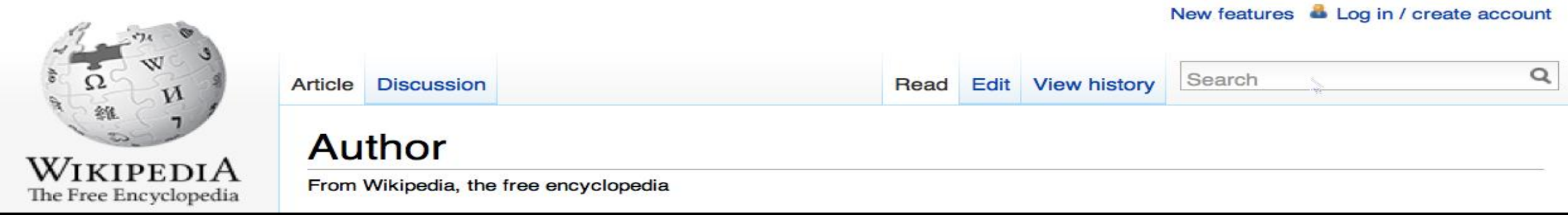
Product
over
Process

Author > Reader

What does it matter who is speaking?

From author to author-function

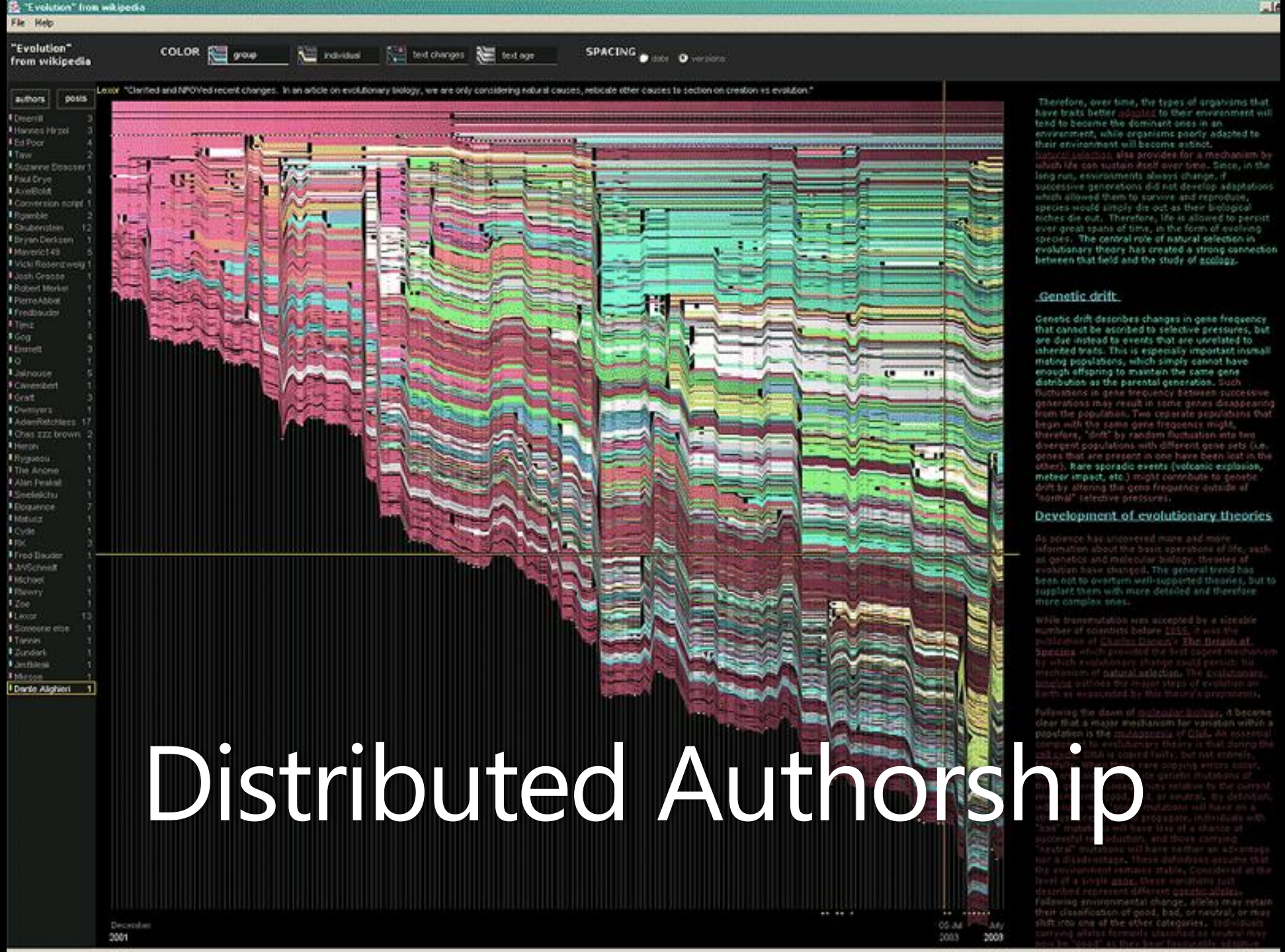
- What are the modes of existence of this discourse?
- Where has it been used, how can it circulate and who can appropriate it for himself?
- What are the places in it where there is room for possible subjects?
- Who can assume these various subject positions?



- collaborative (edit)
- continuous (discussion)
- constant (history)

Reader/Writer/Editor

Serial Collaboration



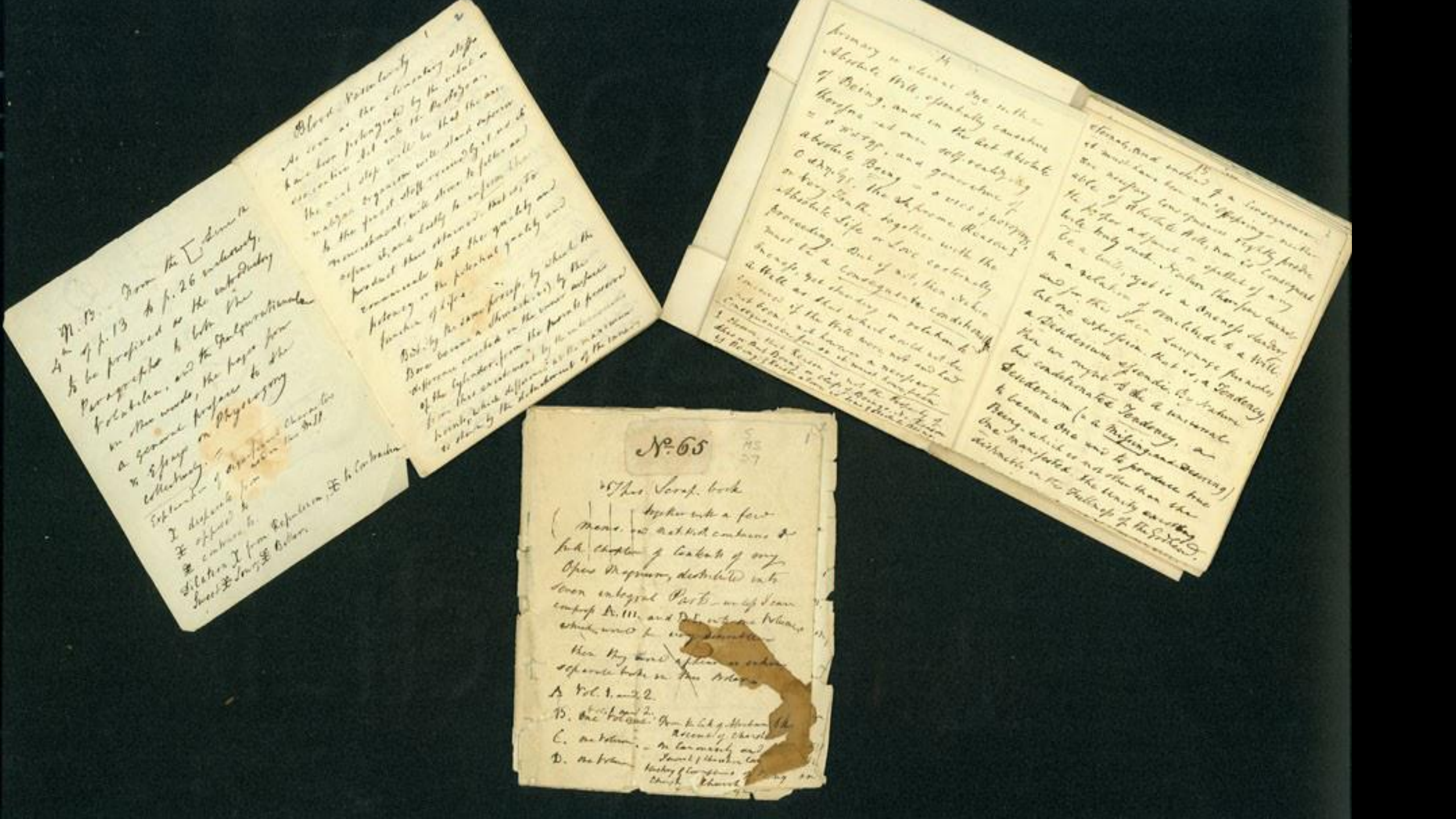
Distributed Authorship

DC/MS/28/10

Fill their Seven Vials with salutary wrath,
To sickly Nature more medicinal
That what soft balm the weeping good man pours
Into the lone despoiled trav'ller's wounds!

Thus from th' Elect, regenerate thro' faith,
Pass the dark Passions and what thirsty Cares
Drink up the spirit and the dim regards
Self-center. Lo they vanish! or acquire
New names, new features — by supernal grace
Enrob'd with Light, and naturaliz'd in Heaven.

As when a Shepherd on a vernal morn
Thro' some thick fog creeps tim'rous with slow foot,
Darkling he fixes on th' immediate road
His downward eye: all else of fairest kind
Hid or deform'd. But lo, the bursting Sun!



Distributed Creation

Crowdsourcing

Anthropology of Free Software Production

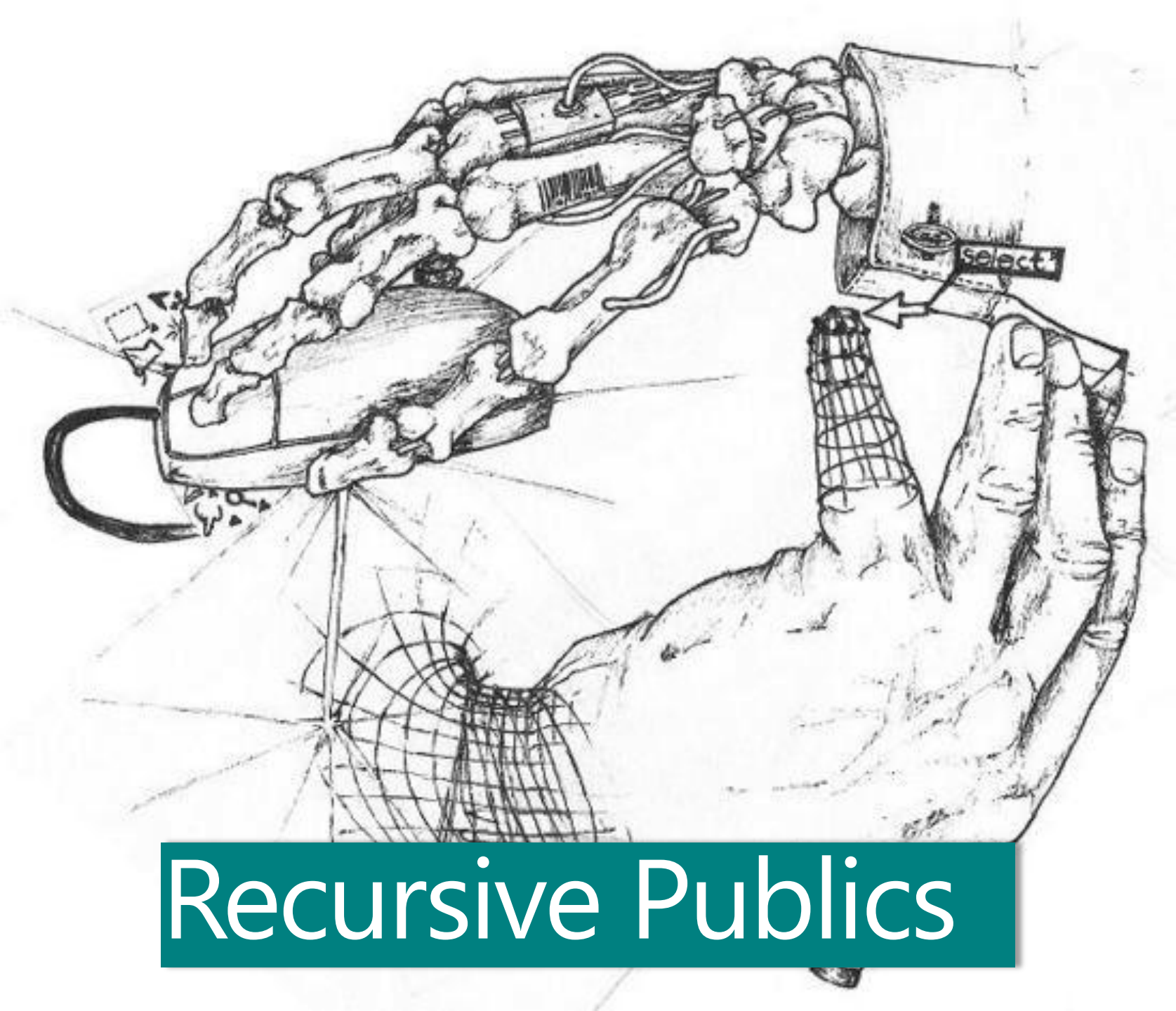
- Highly Organized
- Transparent
- Meritocratic

Massively Distributed Collaboration

Commons-based Peer Production

“Free as in Freedom”

Free Software



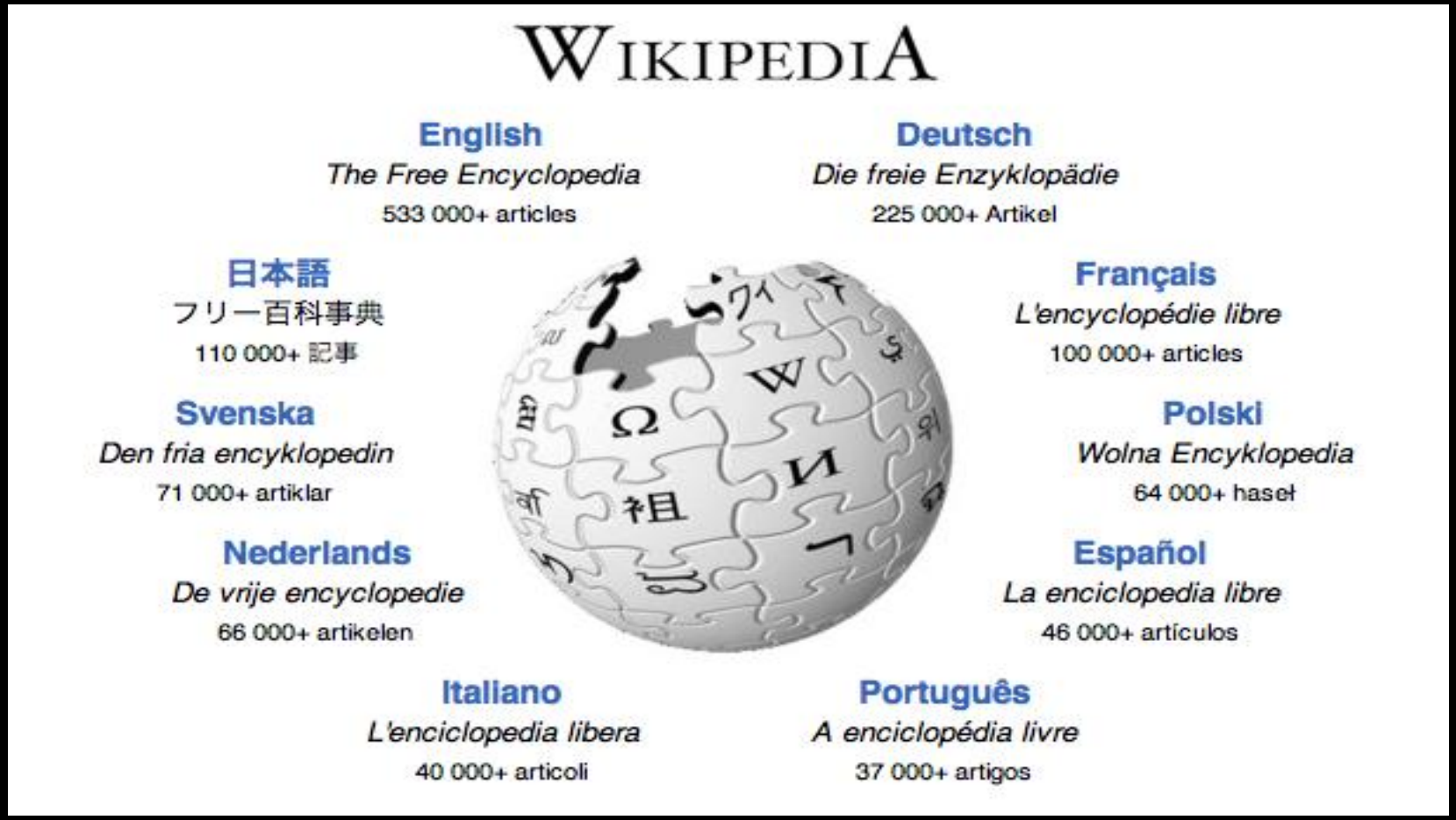
Recursive Publics

Community

Digital Palimpsest

PART THREE

“Planet-Wiki?”
Language,
Translation,
Power.



Currently more than 270 languages.
According to Alexa, as of May 2010 Wikipedia was the 6th most visited site on the planet.

a multilingual free encyclopedia

Wiktionary

[ˈwɪkʃənrɪ] *n.*,

a wiki-based Open Content dictionary

Wikiquote

WIKIMEDIA COMMONS

Wiki Books

Think free. Learn free.

WIKISPECIES

free species directory

Wikiquote

WIKINEWS

WIKIPEDIA

WIKIVERSITY

1 000 000+ articles

№	Language	Language (local)	Wiki	Articles	Total	Edits	Admins	Users	Active Users	Images	Depth
1	English	English	en	**3,318,857**	20,591,387	392,782,085	1,727	12,507,334	149,010	847,764	517
2	German	Deutsch	de	**1,078,924**	3,081,634	78,449,987	288	1,011,248	24,706	192,622	88

100 000+ articles

№	Language	Language (local)	Wiki	Articles	Total	Edits	Admins	Users	Active Users	Images	Depth
3	French	Français	fr	**957,108**	3,887,654	57,520,684	188	848,448	16,276	40,019	139
4	Polish	Polski	pl	**706,166**	1,263,655	23,093,164	164	368,878	6,048	1,176	11
5	Italian	Italiano	it	**696,129**	2,162,567	36,112,756	98	521,356	8,829	77,805	74
6	Japanese	日本語	ja	**682,165**	1,787,921	33,085,529	64	426,114	11,783	74,855	49
7	Spanish	Español	es	**606,773**	2,748,813	40,335,195	138	1,508,207	16,741	0	183
8	Dutch	Nederlands	nl	**605,927**	1,461,600	21,542,138	68	319,918	5,333	18	29
9	Portuguese	Português	pt	**585,218**	2,276,410	20,878,532	38	741,932	6,482	0	77
10	Russian	Русский	ru	**546,735**	1,950,805	26,572,110	87	499,431	12,829	101,264	90
11	Swedish	Svenska	sv	**358,898**	942,021	12,210,704	103	177,694	3,551	0	34
12	Chinese	中文	zh	**311,808**	1,004,649	13,513,399	79	844,792	5,770	28,288	66
13	Norwegian (Bokmål)	Norsk (Bokmål)	no	**262,353**	637,102	7,601,313	64	163,064	2,284	628	24
14	Catalan	Català	ca	**246,148**	590,965	5,523,759	23	61,658	1,528	5,347	18
15	Finnish	Suomi	fi	**240,705**	649,188	8,827,607	49	150,260	2,263	24,134	39
16	Ukrainian	Українська	uk	**211,420**	657,435	4,498,804	20	71,588	1,627	35,124	30
17	Czech	Čeština	cs	**164,915**	427,758	5,652,158	25	114,941	2,181	178	34
18	Hungarian	Magyar	hu	**163,175**	541,727	8,009,882	35	140,412	2,442	34,977	80
19	Turkish	Türkçe	tr	**145,285**	710,793	8,218,727	23	289,641	2,351	23,849	175
20	Romanian	Română	ro	**145,075**	584,659	4,286,536	25	146,561	1,535	41,497	67
21	Korean	한국어	ko	**137,738**	399,548	5,597,803	24	116,889	2,178	7,763	51
22	Esperanto	Esperanto	eo	**130,307**	288,597	2,928,596	18	36,784	477	11,970	15
23	Danish	Dansk	da	**130,055**	345,129	4,244,169	40	105,708	1,260	5	34
24	Arabic	العربية	ar	**127,516**	713,325	6,558,562	23	288,070	2,129	7,897	194
25	Indonesian	Bahasa Indonesia	id	**126,063**	447,132	3,861,738	14	180,394	1,517	20,111	56

30 languages have more than 100,000 entries

61 languages have more than 10,000 entries

96 languages have more than 1000 entries

Meta:Babylon

From Meta, a Wikimedia project coordination wiki

NEW BABYLON

the Wikimedia translators' portal/noticeboard

English: The **Babylon** is the Meta translations portal and noticeboard. For general discussions about translations see the talk page.

+/-

Language select: en Select Show all

Go to meta index

Getting started

- Introduction
- What languages do you speak?
- Translation FAQ
- Editing tutorial
- Work together

Submission help

- Foundation-level issues
  - Read more...
- Submit your own request!

Transcom

- About us · Contact us

Translation requests

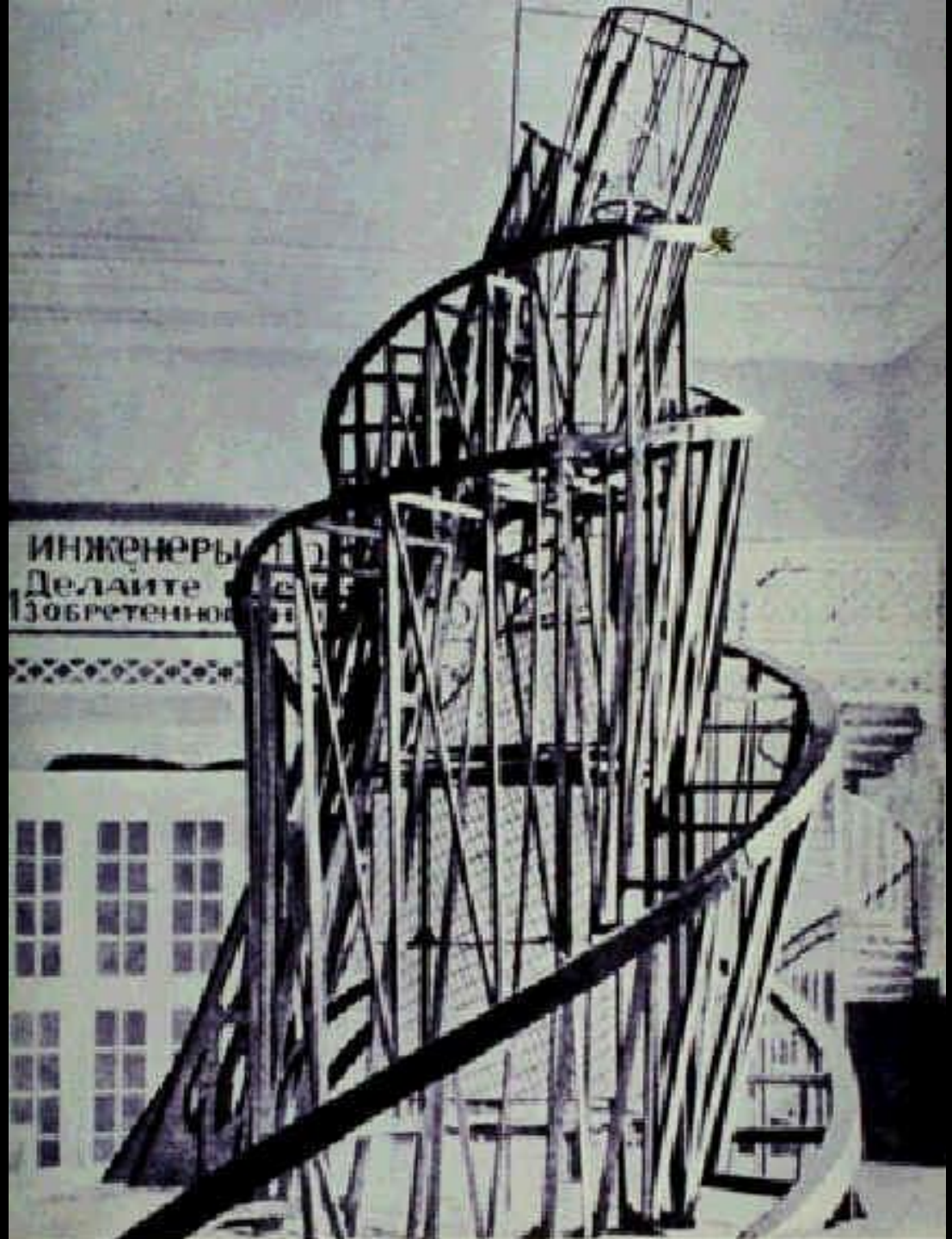
Automagical page listing all translation requests on meta: go there for request-related links.

Direct links:

- Urgent needs
- Need to copyedit/proofread
- Requests from communities
- MediaWiki Localization

Communications

- **Meta talk:Babylon**
  If you have questions, doubts, proposals and anything else about translations you want to ask, write at this page.
- translators-l
  The official Wikimedia translators' mailing list. Sign up!
- #wikimedia-translators
  The official translator's IRC channel. Visit us when you need help, just want to chat, or for


ИНЖЕНЕРЫ
Делайте
Изобретения

The End of History ?

Sources for translation are languages of power

Distributed outwards from the Anglosphere.

Natural Language

- humans are distinguished as a species by their advanced capacity to use language
- language is the means for connecting the internal world of the organism to the external world of natural and social phenomena

"Universal Grammar"
Humans have innate capacity for grammar.
Noam Chomsky, *Logical Structure of Linguistic Theory* (1955)

For a long time, the idea that language might shape thought was considered at best untestable and more often simply wrong. Research in my labs at Stanford University and at MIT has helped reopen this question. We have collected data around the world: from China, Greece, Chile, Indonesia, Russia, and Aboriginal Australia. What we have learned is that people who speak different languages do indeed think differently and that even flukes of grammar can profoundly affect how we see the world. Language is a uniquely human gift, central to our experience of being human. Appreciating its role in constructing our mental lives brings us one step closer to understanding the very nature of humanity.

HOW DOES OUR LANGUAGE SHAPE THE WAY WE THINK? *[6.12.09]*
By Lera Boroditsky

LERA BORODITSKY is an assistant professor of psychology, neuroscience, and symbolic systems at Stanford University, who looks at how the languages we speak shape the way we think.

Lera Boroditsky's *Edge* Bio Page

Sapir-Whorf Hypothesis

- theorizes that thoughts and behavior constructed and influenced by language
- linguistic relativity: “the variance of cultural concepts and categories in different languages affect the cognitive classification of the experienced world in such a way that speakers of different languages *think and behave differently because of it*"
- languages are “arbitrary systems of symbolism”

Linguistic Imperialism?

Translation is not a faithful act of reproduction

Translation is interpretive

What kinds of communities?

The Internet and the politics of development

The Green Revolution

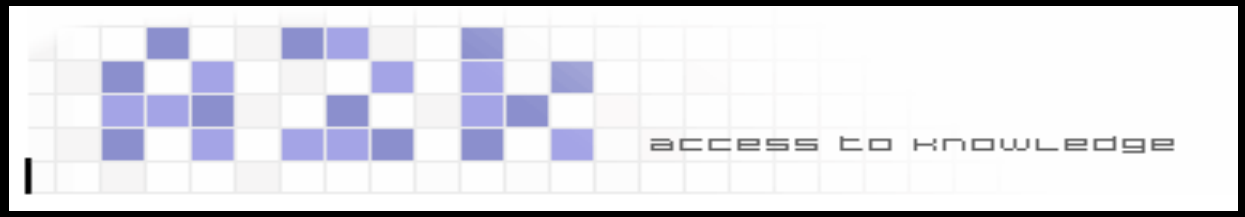
Top Down

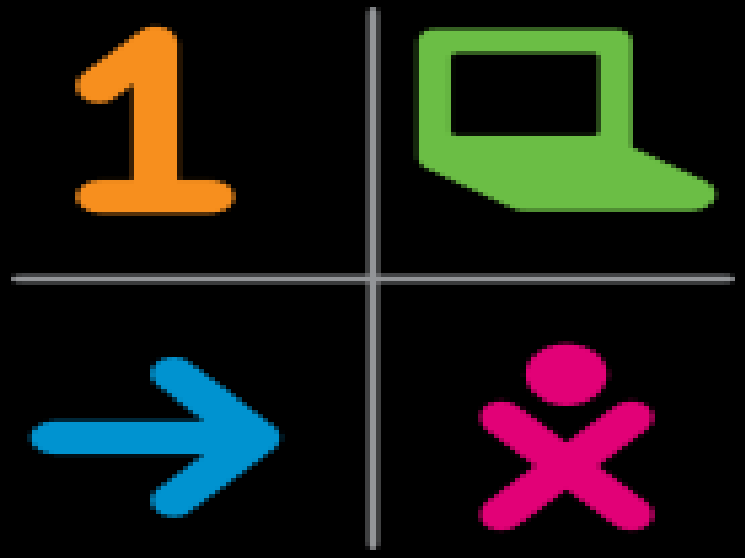
Universalist and Totalizing

Techno-deterministic

Unintended Consequences

one laptop per child


access to knowledge


one laptop per child

...to protect and enhance [expand] access to knowledge, and to facilitate the transfer of technology to developing countries. (A2K Treaty Draft 9 May 2005)

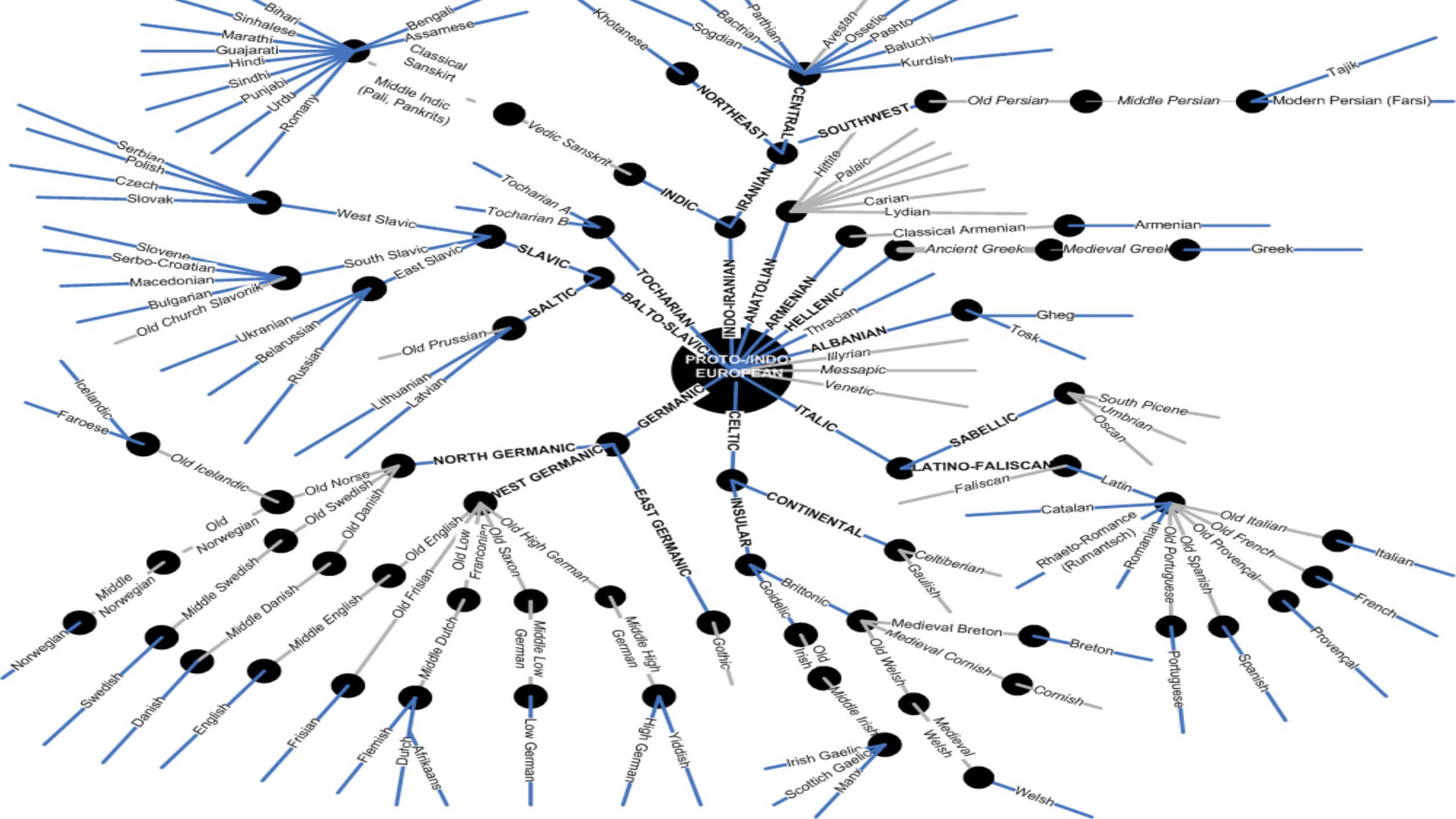
Freedom

- Technological threshold conditions enable greater practical human agency

- Individual action, both commercial and noncommercial, and social sharing & exchange emerge as major modalities of economic production

- Greater individual autonomy, an appreciably more participatory public sphere, and newly emerging practices of a more participatory and critically self- reflective culture

information and access automatically convert into empowerment


PROTO-/INDO-EUROPEAN
INDO-IRANIAN
ANATOLIAN
ARMENIAN
HELLENIC
ALBANIAN
TOCHARIAN
BALTO-SLAVIC
SLAVIC
BALTIC
GERMANIC
CELTIC
ITALIC
INDIC
IRANIAN
NORTHEAST
CENTRAL
SOUTHWEST
NORTH GERMANIC
WEST GERMANIC
EAST GERMANIC
INSULAR
CONTINENTAL
SABELLIC
LATINO-FALISCAN
Bihari
Sinhalese
Marathi
Guajarati
Hindi
Sindhi
Punjabi
Urdu
Romany
Bengali
Assamese
Classical Sanskrit
Middle Indic (Pali, Pankrits)
Vedic Sanskrit
Khotanese
Sogdian
Bactrian
Parthian
Avestan
Ossetic
Pashto
Baluchi
Kurdish
Old Persian
Middle Persian
Modern Persian (Farsi)
Tajik
Hittite
Palaic
Carian
Lydian
Classical Armenian
Armenian
Ancient Greek
Medieval Greek
Greek
Thracian
Gheg
Tosk
Illyrian
Messapic
Venetic
Tocharian A
Tocharian B
Serbian
Polish
Czech
Slovak
West Slavic
Slovene
Serbo-Croatian
Macedonian
Bulgarian
Old Church Slavonic
South Slavic
East Slavic
Ukranian
Belarussian
Russian
Old Prussian
Lithuanian
Latvian
Icelandic
Faroese
Old Icelandic
Old Norse
Old Norwegian
Middle Norwegian
Norwegian
Old Swedish
Middle Swedish
Swedish
Old Danish
Middle Danish
Danish
Old English
Middle English
English
Old Frisian
Frisian
Old Low Franconian
Middle Dutch
Flemish
Dutch
Afrikaans
Old Saxon
Middle Low German
Low German
Old High German
Middle High German
High German
Yiddish
Gothic
South Picene
Umbrian
Oscan
Faliscan
Latin
Catalan
Rhaeto-Romance (Rumantsch)
Romanian
Old Italian
Old French
Old Provençal
Old Portuguese
Old Spanish
Italian
French
Provençal
Portuguese
Spanish
Celtiberian
Gaulish
Brittonic
Goidelic
Medieval Breton
Breton
Old Irish
Middle Irish
Irish Gaelic
Scottish Gaelic
Manx
Old Welsh
Medieval Cornish
Cornish
Medieval Welsh
Welsh

irreducible diversity of language

design heterogeneity

Networked Cultures of Information and Identities

- The humanities, like all disciplines, must recognize the shift to digitality. In doing so, it must begin to explore the intermingling between natural and machine languages, between analog and binary code.

- The forms of interaction produced by the interactive web necessitate new ways of understanding modern concepts such as authorship, identity, property, credit, expertise and knowledge.

Beyond Development

- Despite networked technologies that long predate the web, the liberal humanist subject continues to serve as the locus of identity, despite all evidence to the contrary.

- The proliferation of communications technologies threatens to re-enforce rather than undermine long-standing histories of colonialism and domination. Like economic imperialism, cultural imperialism displaces heterogeneity and complexity with mechanistic, undifferentiated, essentializing approaches. These approaches extend not just to technology, but information, knowledge and what constitutes ‘understanding’.

Microsoft®

© 2010 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.
The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.
MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.

Microsoft® Research

Faculty Summit 2010