

Governing the virtual commons?

Questions around and from Wikipedia

Giovan Francesco Lanzara
University of Bologna and ISIG, London School of Economics

&
Aleksi Aaltonen
ISIG, London School of Economics

ICT in the Contemporary World, LSE
November 17th, 2009

The focus of our inquiry

- Emergence of large-scale Internet-mediated social systems
 - Supported by underlying ICT infrastructures
 - Exchange, communication, production, organizing
 - The Internet has revealed (and enacted) a hidden potential for social interaction and collective action buried in the texture of societies
- Commons-based peer production (Benkler 2006)
 - Neither market nor firm
 - It epitomizes a new kind of mediation between individual contributions and collective engagements (Kallinikos 2009)
 - It directly converts online sociality into productive force
 - Low transactions costs
- Governance of distributed knowledge systems
 - How are such systems ruled?
 - Where and how should resources be allocated in the design of collaborative knowledge systems?

- **What is a commons?**
 - *“a natural or man-made resource system that is sufficiently large as to make it costly (but not impossible) to exclude potential beneficiaries from obtaining benefits from its use” (Ostrom 1990, Ostrom 2000)*
 - *“a large number of distributed interacting people who attend a common production task sharing collective resources on a decentralized, nonproprietary model” (Benkler 2006)*
- **How is commons-based peer production governed over the Internet?**

The case of Wikipedia: who writes it? who/what runs it?

Why Wikipedia?

- an existence proof and a model for distributed cognition
- a concrete embodiment of distributed collective intelligence (swarm intelligence)
- an evolving entity: **not today what it was yesterday; not tomorrow what it is today**
- The thing itself and its history are all there, available for inspection!
- *Much used, much talked about and studied, but little understood*

Questioning the Wikipedians' self-image

- “The free encyclopedia that anyone can edit”
- “A community of free editors abiding to hundreds of self-enforced rules”
- Discourse-centric model of cooperation based on social norms
- Whose image is this? Espoused theory and actual practice

- Who does what? Who contributes what?
- How has it evolved throughout the years?
- How does governance happen in Wikipedia?
 - “*self-conscious social-norms-based dedication to objective writing*” (Benkler 2006)
 - “*...some combination of technical architecture, social norms, legal rules, and a technically backed hierarchy that is validated by social norms*” (Benkler 2006:104)
- **Are community-based mechanisms adequate to govern such a large armada of people attending to the same complex task and stepping in and out?**

current research

- Aggregate demographics: edits of articles, users
- Issues of quality assurance and reliability; NPOV
- Some studies on coordination and on workflow
(Kittur et al 2007; Viègas et al. 2007)
- Some studies on decentralization and policy
(Forte et al 2009; Beschastnikh et al 2008)

Little understanding of the organizational dimension and the complexity of practices in Wikipedia

Unless the evolving organization and regulation of the production work is clarified Wikipedia will not be properly understood and questions about governance will not be satisfactorily answered

Wikipedia is an organizational phenomenon

- **Nature of the task; task complexity**
 - not just writing articles and stacking them up in a digital storehouse, but collectively designing and making an encyclopedia
 - Not a list or inventory of items, but a system of categories
 - A web of differentiated and interconnected practices
- **Nature of control possible over human resources**
 - what control mechanisms for such a dispersed and fluid system?
 - not just control of deviant behaviour, but management of a complex production system
 - Low transactions costs *versus* proliferation of rules
- *Answers depend on the metrics used to measure contributions*

Who writes Wikipedia?

Different metrics, different stories

a) **If the number of edits are counted, we get the power/rank distribution (Zipf law, 1949)**

- 50% of all the edits are done by less than 1% of users
- 73,4% of all the edits are done by 2% of the users (around 1400)
- The remaining 25% or so of the edits are introduced by 98% of the users, a large crowd who seems to contribute almost nil
- Jimbo Wales' story of the 'gang of 500': a small group of dedicated volunteers does most of the work, the crowd adds small fixes

(the figures date from the end of 2006:

<http://www.aaronsw.com/weblog/whowriteswikipedia>))

These figures tell us very little about the kind of contributions that have been done; they don't capture what's going on in Wikipedia

Who writes Wikipedia?

b) If the number of letters or words are counted, we get a different story (Schwartz 2006; Kittur & al 2007)

- Editors with many edits contribute little content
- Creation of substantive content is widely distributed among a large mass of people, while the editing work is concentrated with a restricted group of editors (that account for most of the edits)
- Single entries are provided by one or few contributors, then the insider editors come in and fix the text
- From Sept 2002 to Sept 2006 the contribution of elite editors (admins) has declined from 60% to 10%, while at the same time the base of all users has expanded enormously
- There is some significant overlapping between content providers and formatters

(based on limited data base: a few articles)

So, who actually does what in Wikipedia?

- **The expanding intellectual proletariat** of the powerless contributors provides the substantive content, but has no control over the editorial process and the edited product (the wisdom of the crowd)
- **The rising editorial bourgeoisie** leads and controls the editorial process, crafting the encyclopedia as a coherent product and ruling Wikipedia as an organization (the power of the few) (Kittur & al 2007)
- **Voting rules:** eligibility to admin positions is based on number of edits
- *Why has strictly editorial work been concentrated throughout the years?*
- *Why there has been such a dramatic shift of the workload to the 'common' users?*
- *Is this an accurate picture?*

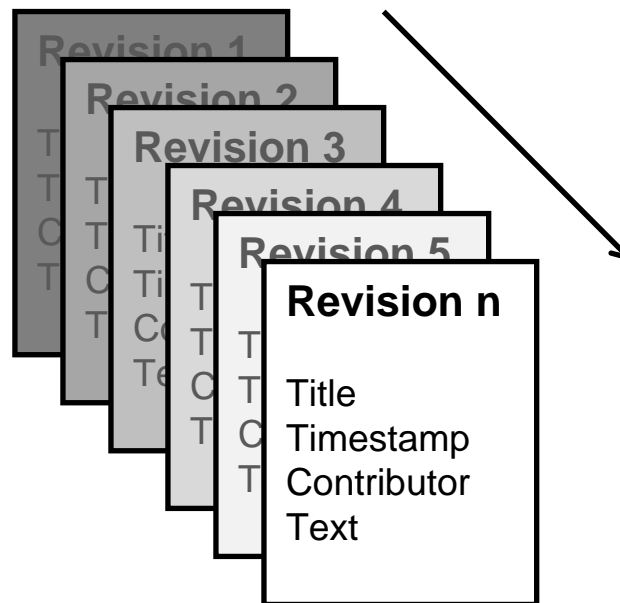
Evolving coordination requirements

- When contributors are still a reasonably small number, coordination and quality standards can be provided by informal communication at low cost (explicit coordination)
- When the number of contributors increases, the costs of coordinating them also increase, and ‘formal structure’ is required to curb costs *ex ante*
- The only way to curb coordination and overhead costs and improve quality is to concentrate editing into a few hands: after a certain threshold it is simply impossible to add more editors to work on an article, it will harm quality (Brooks 1975)
- So, basically, in order to keep the system open, structure is designed into the editorial process, some critical functions are concentrated, and access to specific actions and functions is progressively restricted
- Endogenous evolution: the governance structure has evolved as a response to the shifting requirements generated by the very evolution of Wikipedia as a product and a social system.

What we don't know yet

- No clear idea of the evolving task and role structure
- No clear idea of the evolution of rules and their influence
- No clear idea of the shifting texture of the editing and organizing practices
- No clear idea of how technology works
- **A new research agenda is in order to explore the inner guts of Wikipedia**
 - We need to better understand the technical and organizational architecture of Wikipedia
 - Are there stable groups or clusters of specialized agents taking care of specific tasks?

Wikipedia is made of pages



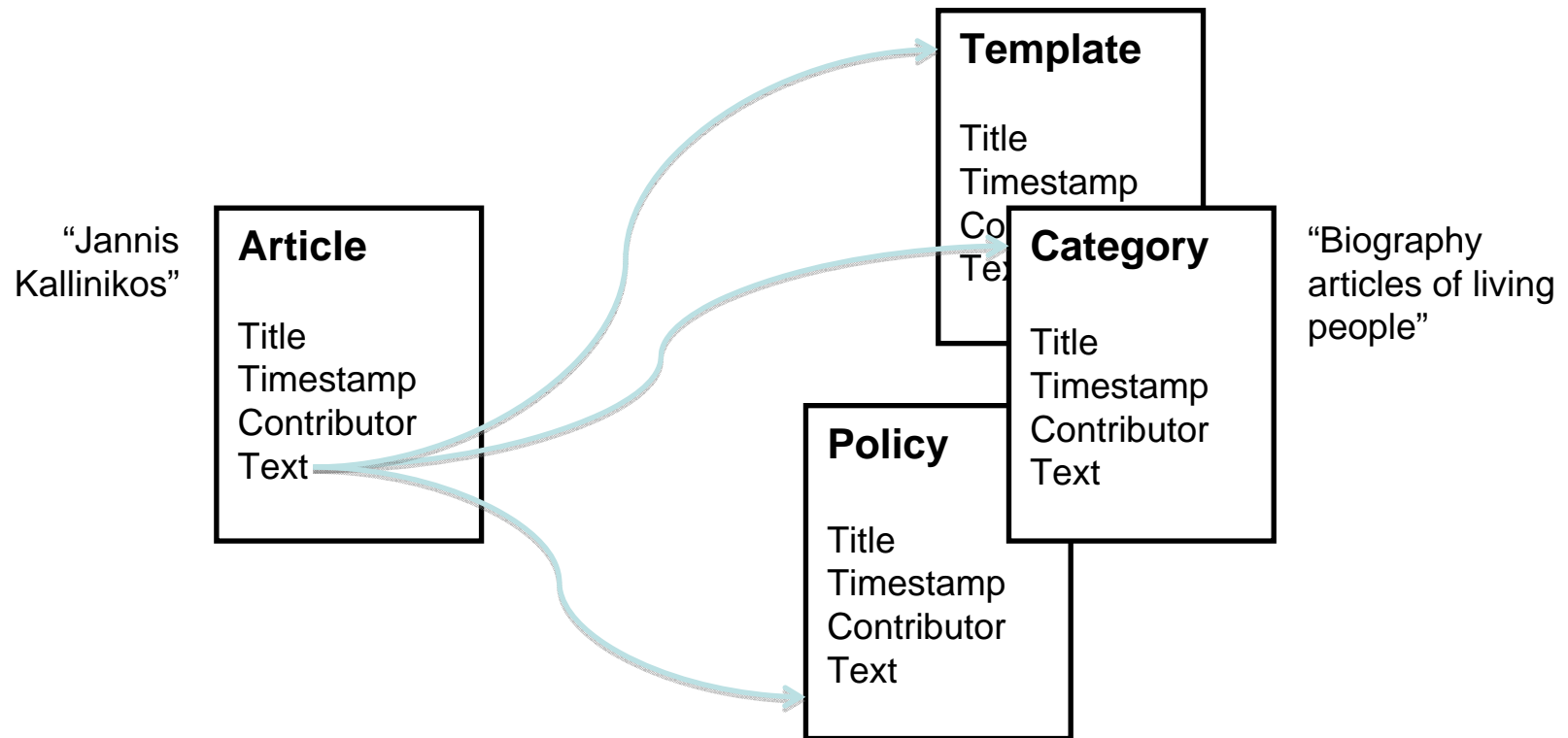
Each modification creates a new revision of the page

Page belongs to a namespace

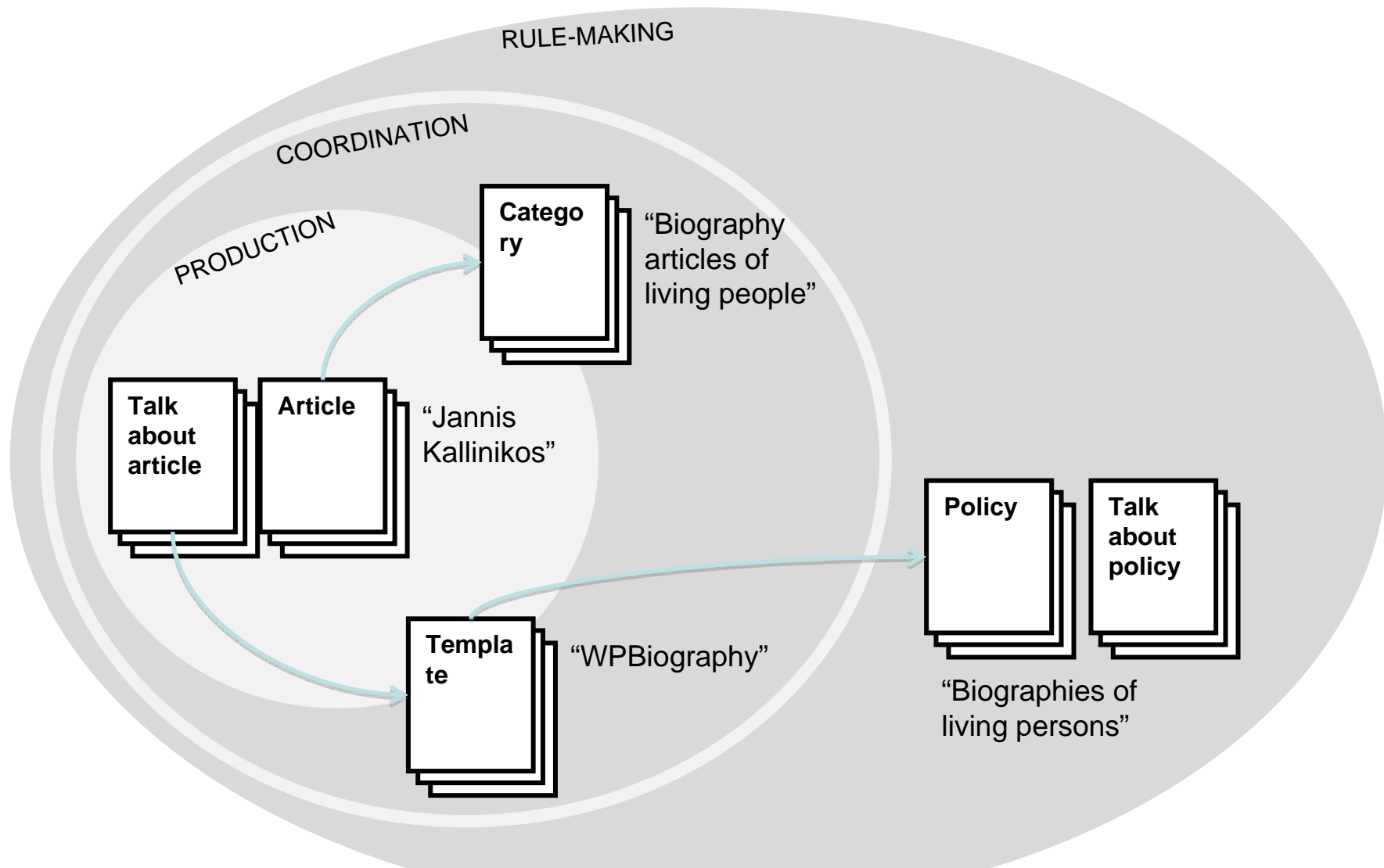
Encyclopedia content	Main	Encyclopedia articles	Communication about particular content
	Talk	Talk about encyclopedia articles	
	User	Registered users	
	User talk	Talk about users' pages	
Policy	Wikipedia	Pages about Wikipedia project itself	Communication about particular policy
	Wikipedia talk	Talk about Wikipedia project pages	
	Template	Pages used as templates for other pages	
	Template talk	Talk about templates	
	Category	Pages that describe categories of pages	
	Category talk	Talk about categories	

Namespaces have different purposes in Wikipedia

Links between different pages



Articles are not just linked to other article pages, but to a variety of categorizing, formatting and administrative devices that are pages themselves.

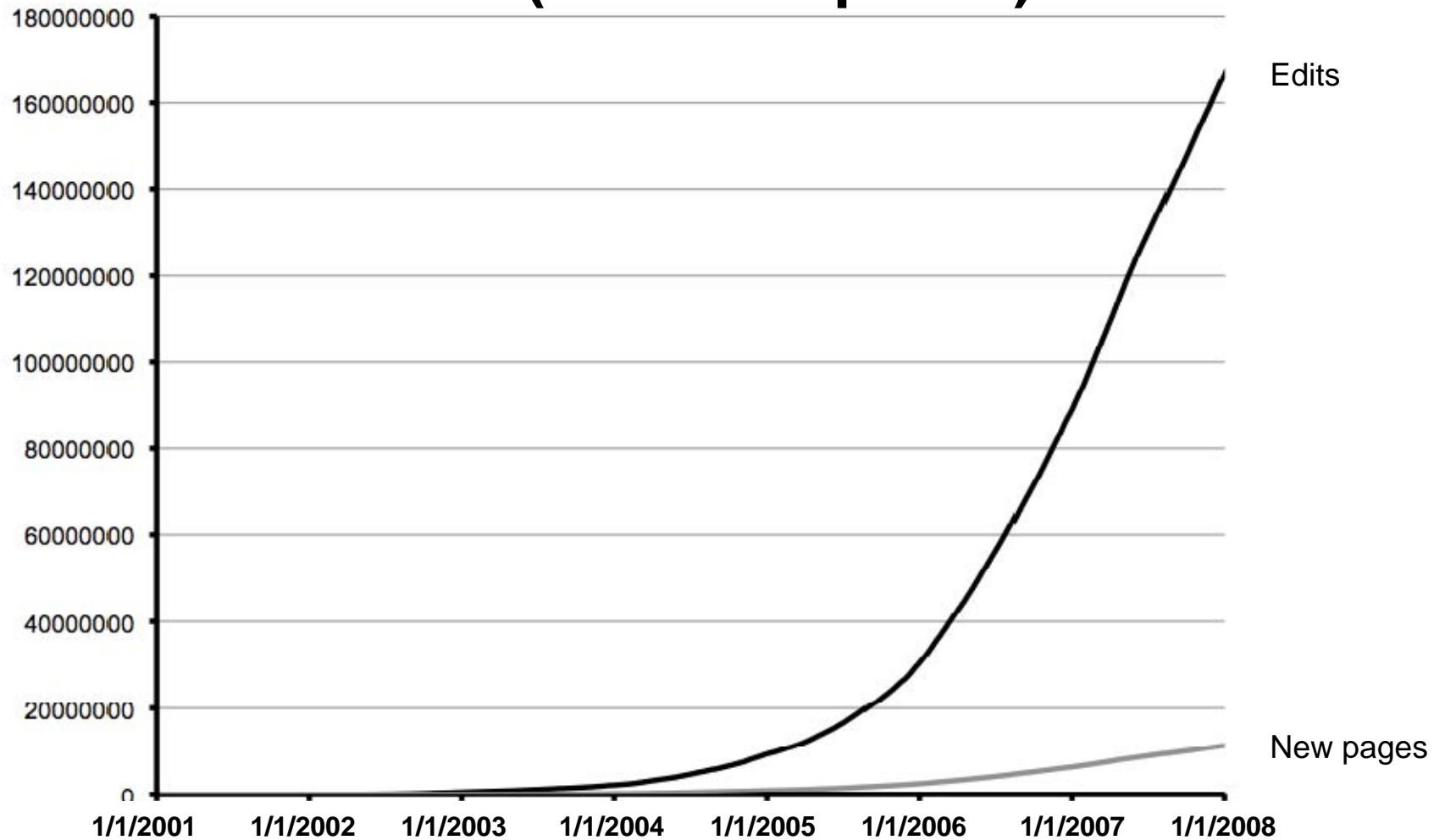


Wikipedia is a complexly embedded self-similar structure. The technology does not only provide a platform for writing articles, but embeds rules about article writing as well as about writing-about-writing, using the very same infrastructure.

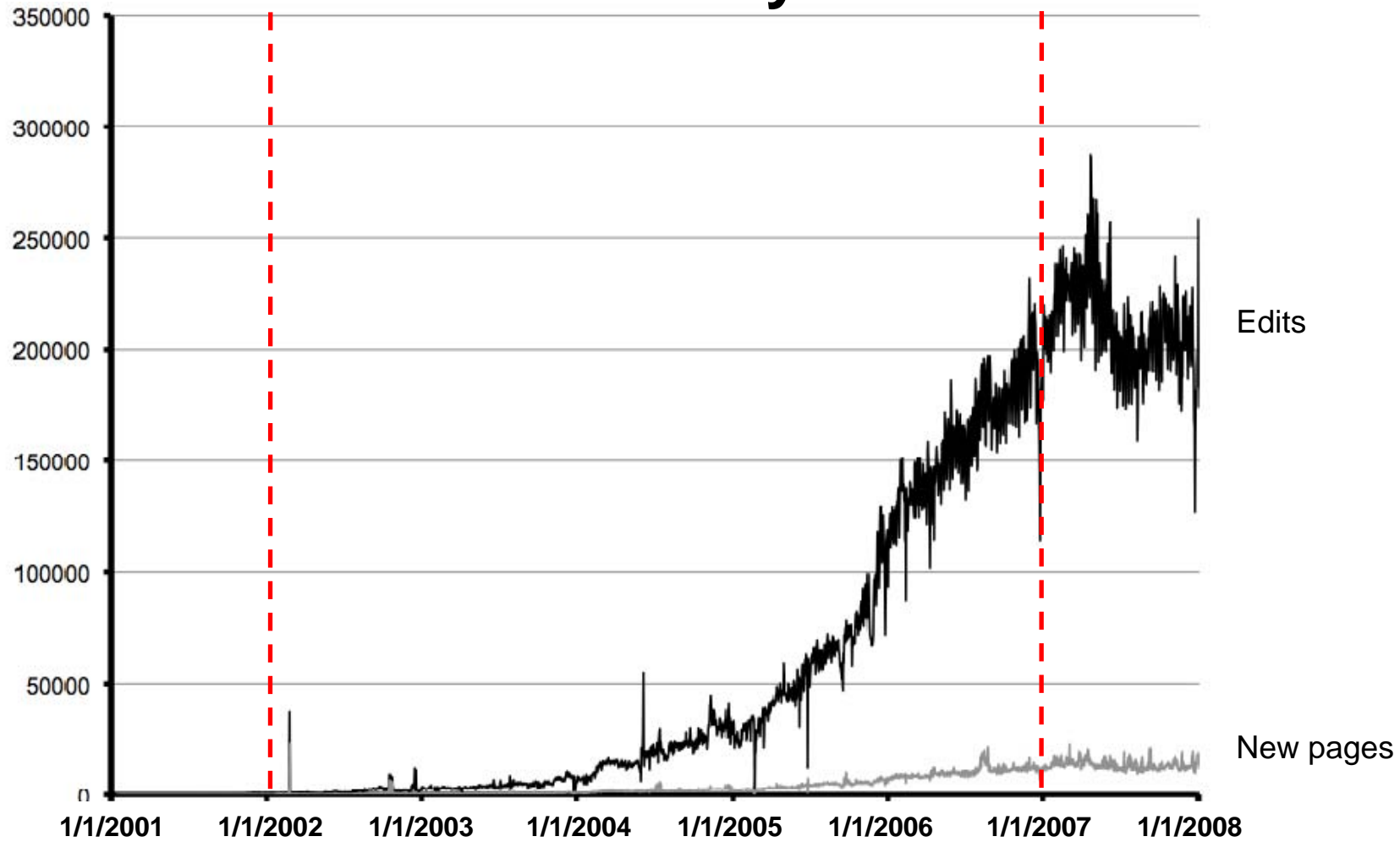
Main analytical steps

- Step 1: Delineating the growth in the different parts of Wikipedia
- Step 2a: Tracing the history of rules and regulatory mechanisms
- Step 2b: Mapping the type of user to the type of edit
- Step 3: Testing the impact of regulatory interventions

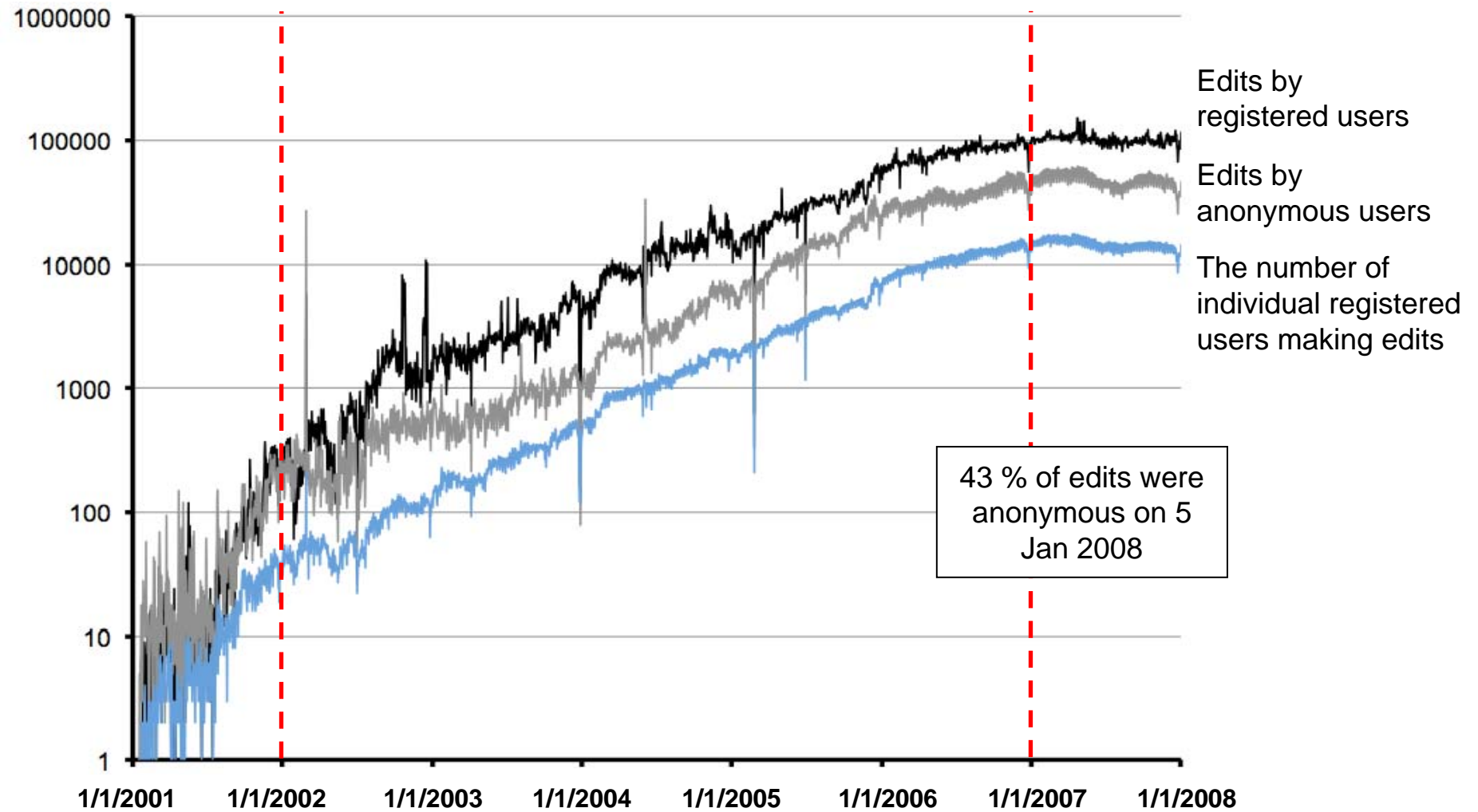
Cumulative edits and new pages (all namespaces)



Edits and new pages (all namespaces) per day

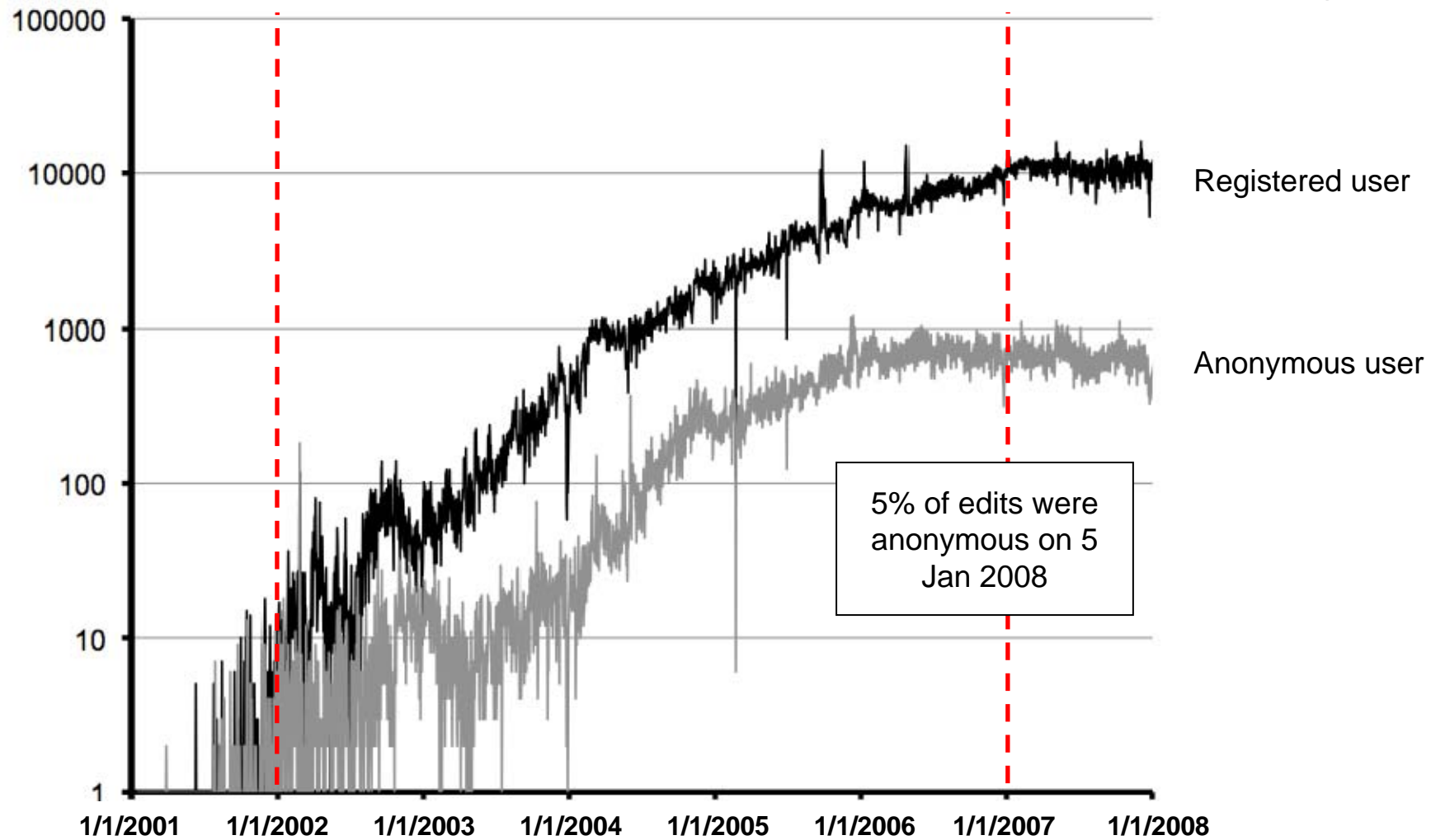


Main namespace (article) activity per day

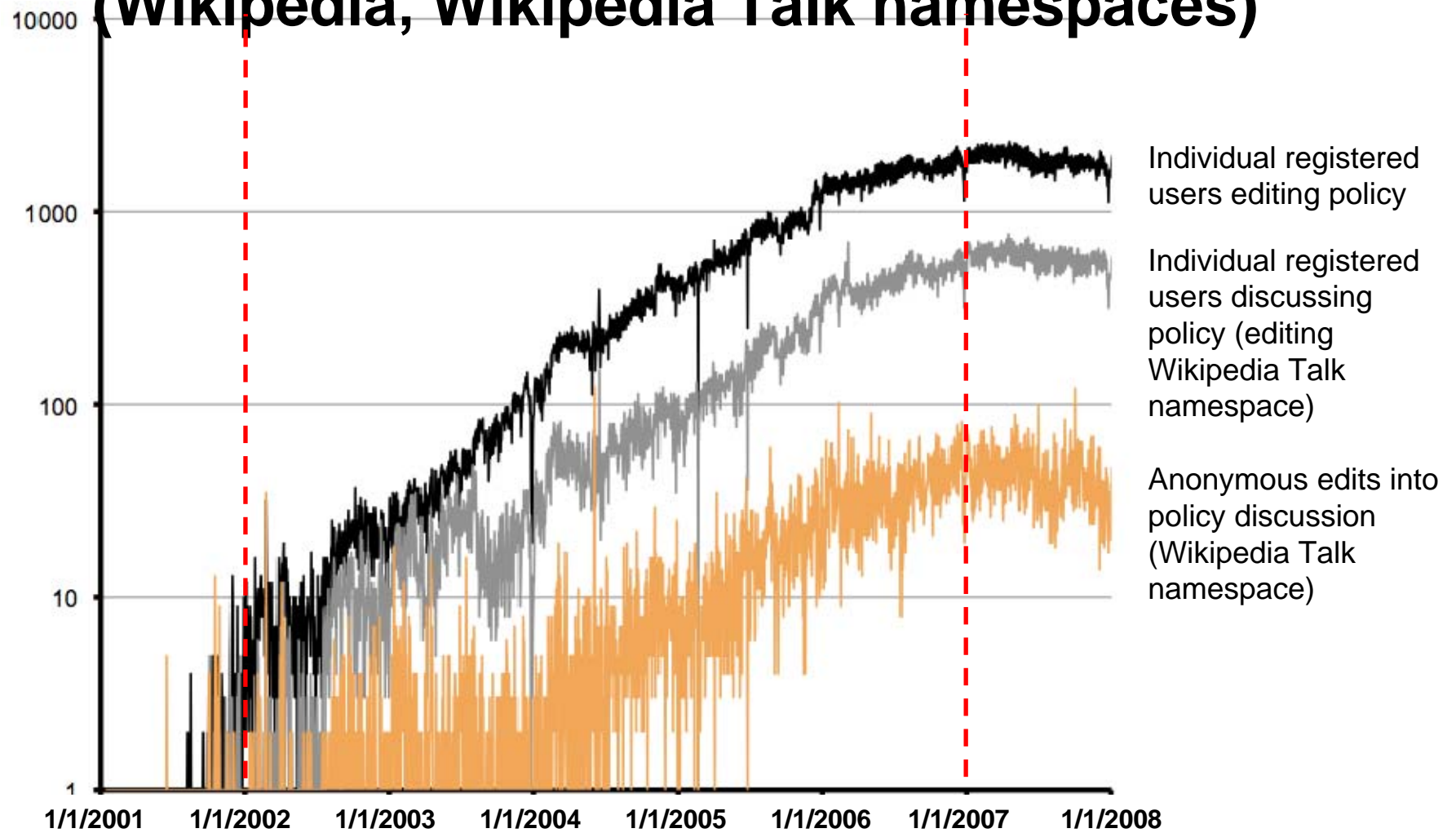


wikipedia/LSE2009

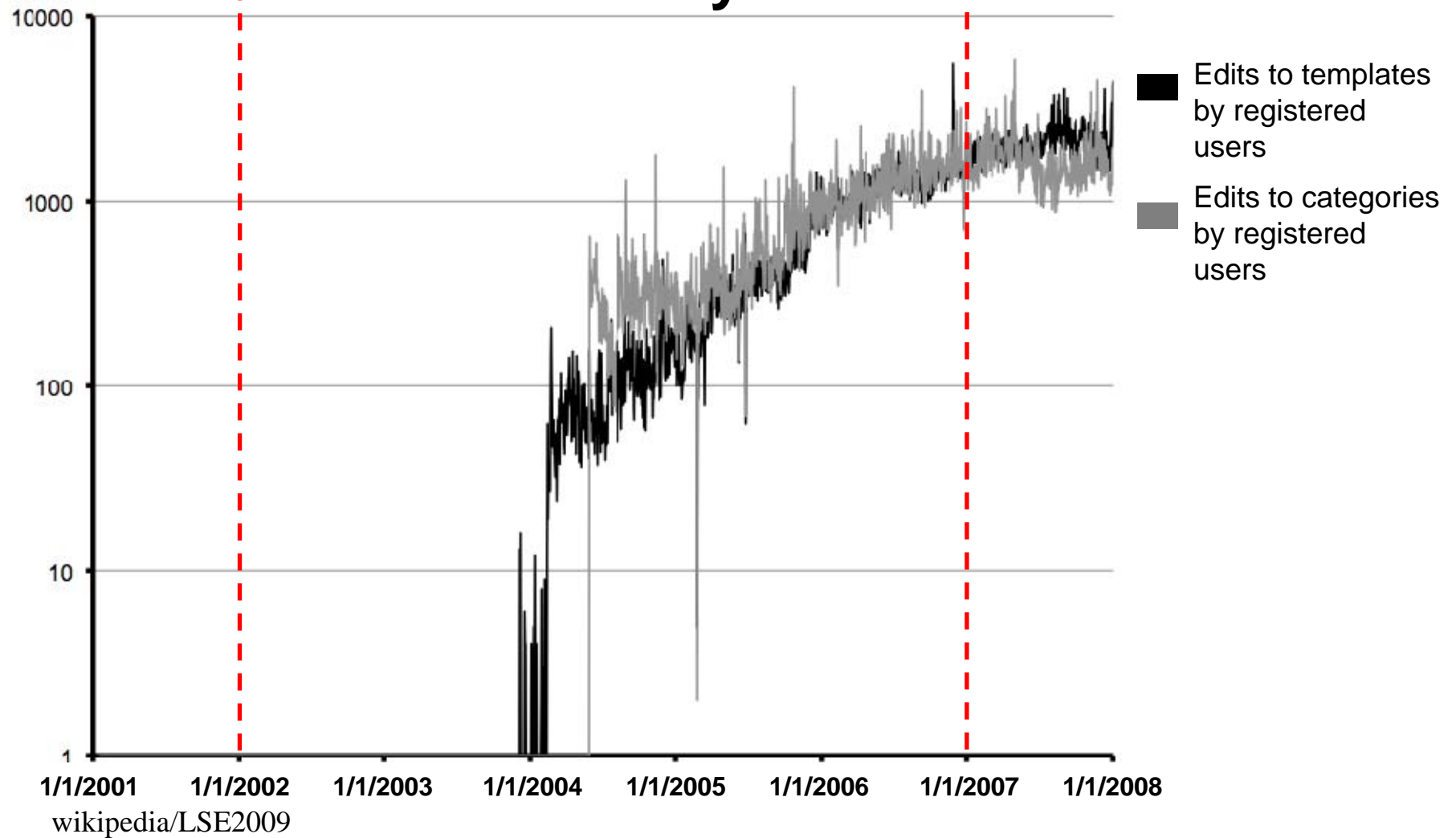
Wikipedia namespace (policy) edits per day



Breadth of participation in policy-making (Wikipedia, Wikipedia Talk namespaces)



Template and Category namespace edits per day



Tracing Wikipedia's evolution

<p>Birth and infancy (Jan 2001 - 2003)</p>	<p>Content production and value creation; community building; critical mass; low coordination costs; emphasis on openness, freedom, community, identity; production first, quality later. Governance based on communitarian ideology. incentives to entry and participation.</p>
<p>Growth (early 2004 - late 2007)</p>	<p>Take off; high increase in low participation users and simultaneous declining proportion of high edit user influence; commons-oriented peer production; issues of quality, reliability, coordination; implementation of new wiki-based tools; production of norms and policies; emergence of an admin hierarchy; rising coordination costs; quality critical for accreditation.</p>
<p>Maturity (2007 - ?)</p>	<p>Increasing task complexity; interdependencies; increasing need for structure; evaluation bodies are created; professionalization of the editing work; career paths; rising coordination costs; more production, editing, formatting standards; structuring and formalizing communication. Stratification of the Wiki society; Governance issues.</p>
<p><i>Decline</i> (in a not too distant future...)</p> <p>wikipedia/LSE2009</p>	<p><i>Shift from production to maintenance and control; increased bureaucratization of the workflow; dispute-resolution mechanisms; formalisms and controls stifle participation; predictable decline of inflow; exit; problems of long tem sustainability of cooperation; entropic pressures.</i></p>

Wikipedia's governance dilemma

How to keep attracting user-generated content while at the same time steering the editing process so that quality, coverage and reliability of articles are maintained and improved

Conflicting pressures and requirements

- Openness vs filtering and monitoring
- Freedom vs control and protection (safeguarding the product)
- Distributed participation vs coordination
- Self-selection vs user registration
- Generalized accessibility vs quality and reliability
- Incentives to participation vs sanctions to sloppyness, vandalism and sabotage

Modes of governance

- Peer-based or self-governance
 - communication, talk, discourse-centric
- By formal structure and rules
 - hierarchical stratification, functional specialization
- By technology
 - Technology ‘mediates’ organizing: but the mediation doesn’t simply transpose organizing isomorphically, it re-frames and re-mediates organizing as a process and as a set of practices
 - As organizations have absorbed society in the course of the XXth century (Perrow 2005), so now technology absorbs organization

ICT as an organizing and governance mechanism

- A technical **platform** or **infrastructure** that facilitates low cost interaction (connectivity)
- An available set of **software devices** that support collective work
- Technology is used to extend **public visibility** into the process
- Software-based devices inscribe interaction and governance functions
- **Complex signaling system that provides orientation**

• Documentation devices	Access devices	Record keeping
• Monitoring	Talk pages	Templates
• Categories	Watchlists
• <i>Notifications</i>	<i>Warnings</i>	<i>Alerts</i>
• <i>Prompts</i>	<i>Things-to-be-done</i>	<i>Do-nots</i>

The signal system is embedded in the production system

Open questions

- What kind of data and metrics do we need to understand how Wikipedia is ruled?
- How can we connect individual users or group of users to sepcific tasks?
- How should resources be allocated in the design and governance of collaborative knowledge systems?
 - Early phase: building tools for power users and improving expert features
 - Mature phase: improving ease of use and effectiveness for novice users; developing structures and procedures supporting a large influx of users
- Generalizability to other Internet-based phenomena? Nothing new?
- What does the digital mediation do to organizing and governance?