

# **Extending the Capabilities of the Internet to the Rural Developing World**

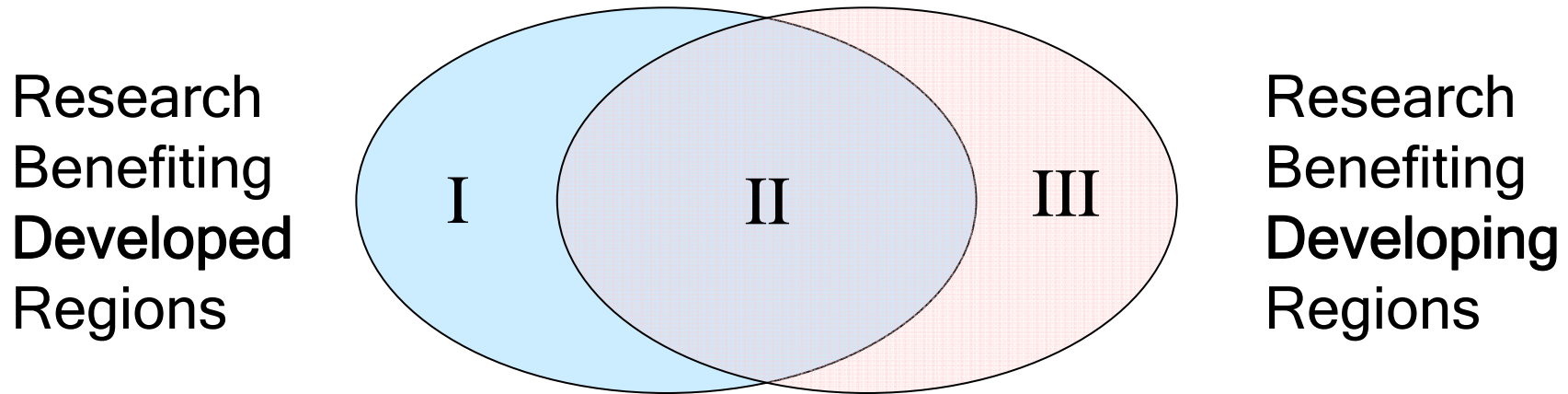
**Bill Thies**

---

Computer Science and Artificial Intelligence Laboratory  
Massachusetts Institute of Technology

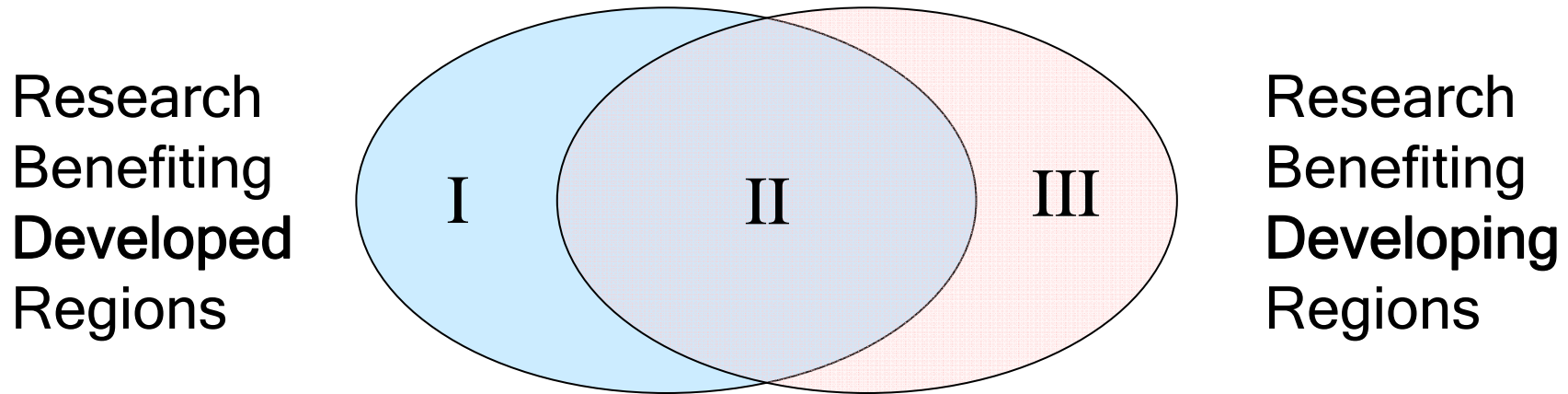
Carnegie Mellon University  
April 25, 2008

# Technology *Research* for Developing Regions



- Most research falls in intersection (II)
  - Invented for developed areas, later applied in developing areas
  - Examples: computers, cell phones, Internet, ...
  - Many people are working in this area

# Technology *Research* for Developing Regions



- How to target developing regions (III)?
  1. Identify **trend, constraint** or **opportunity** that is present in developing world, but not in developed world
  2. Invent technology to **exploit** or **compensate** for the trend
  3. Result would **not have been invented** in developed world
- Strive for deep, novel, long-term research innovations
  - Fewer people working in this area → large potential impact

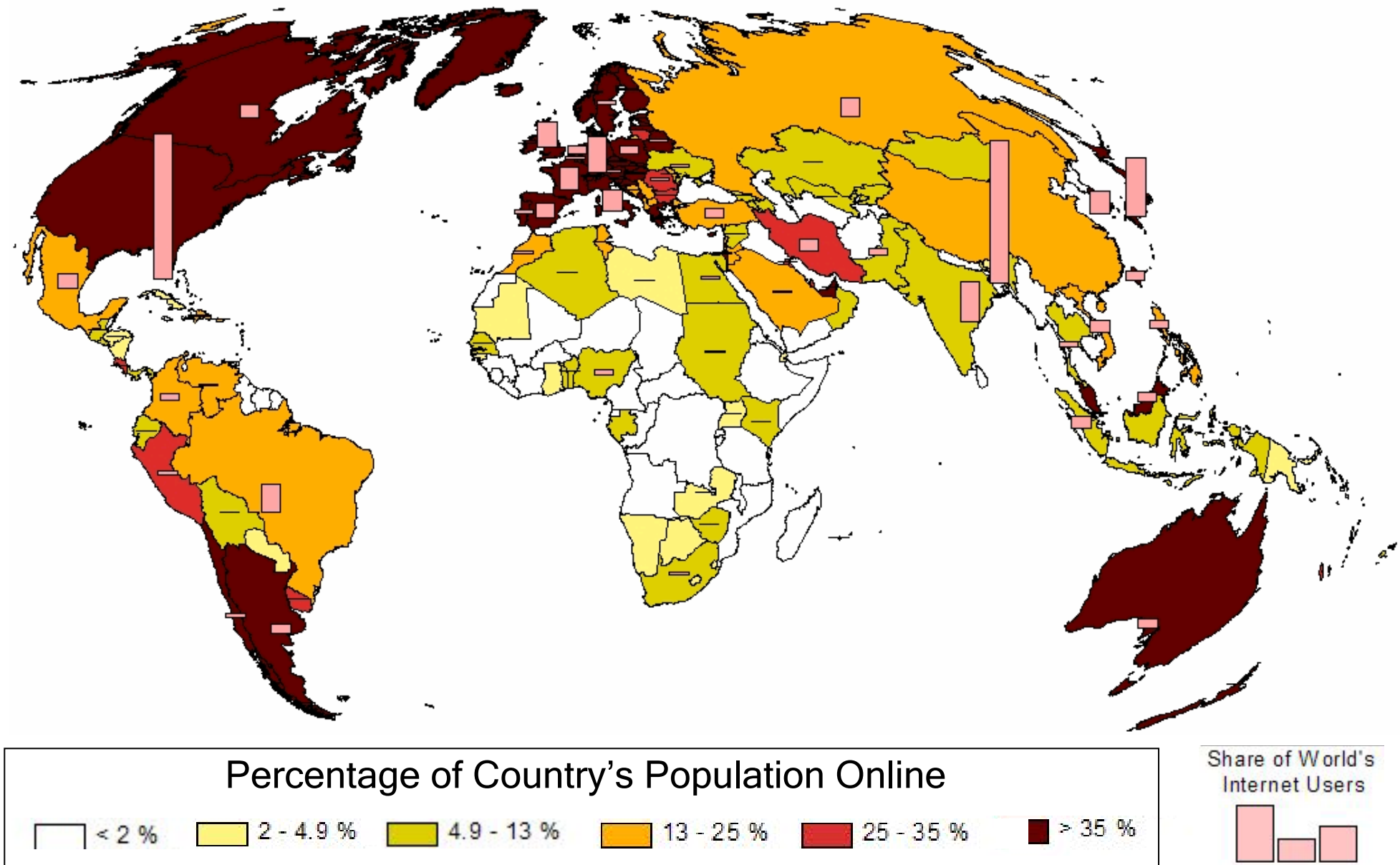
# In This Talk: Three Directions

1. Retrieving information
2. Publishing information
3. Collecting medical data

# In This Talk: Three Directions

1. Retrieving information
2. Publishing information
3. Collecting medical data

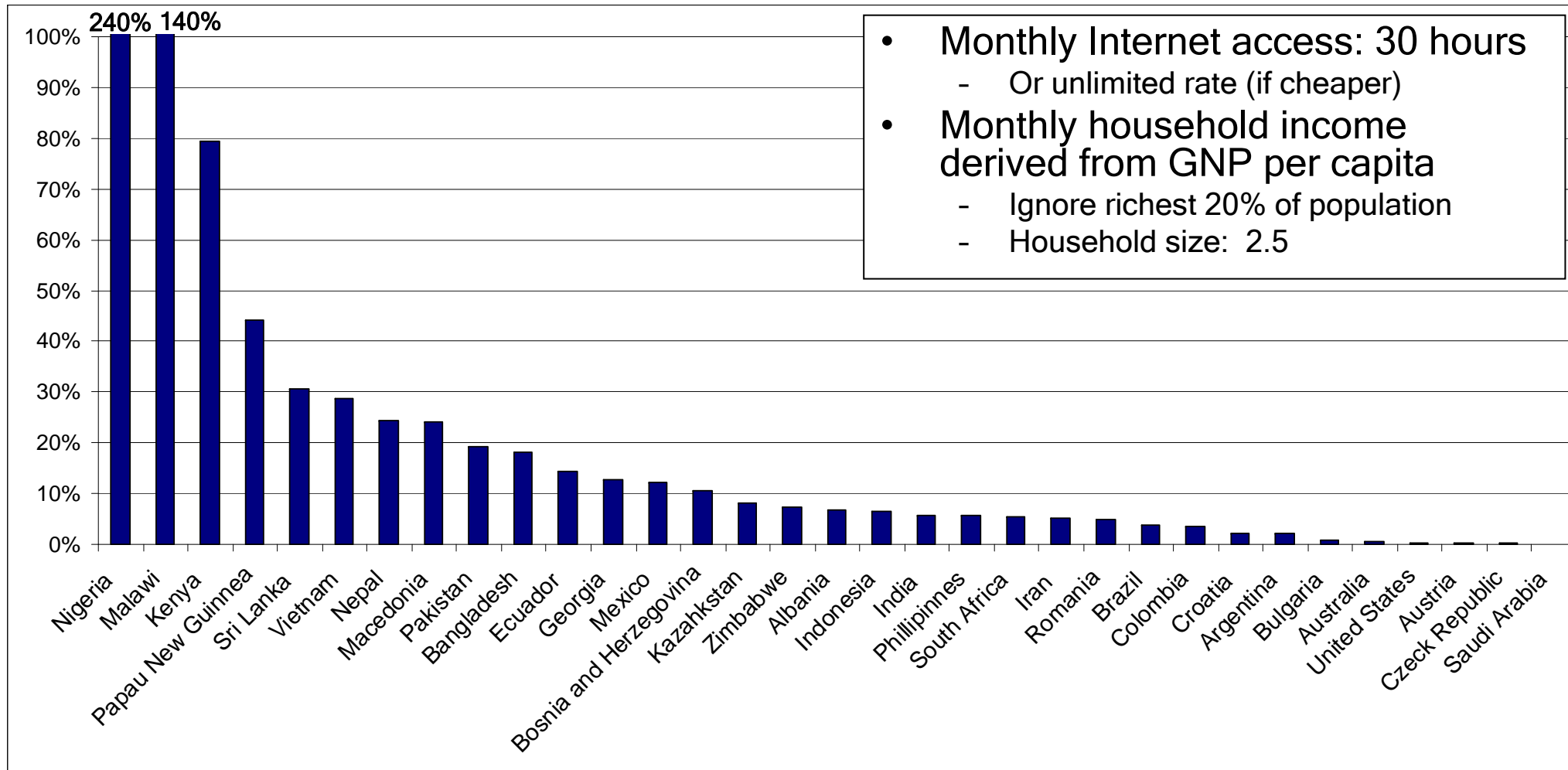
# Internet Users Worldwide



# Barriers to Internet Access

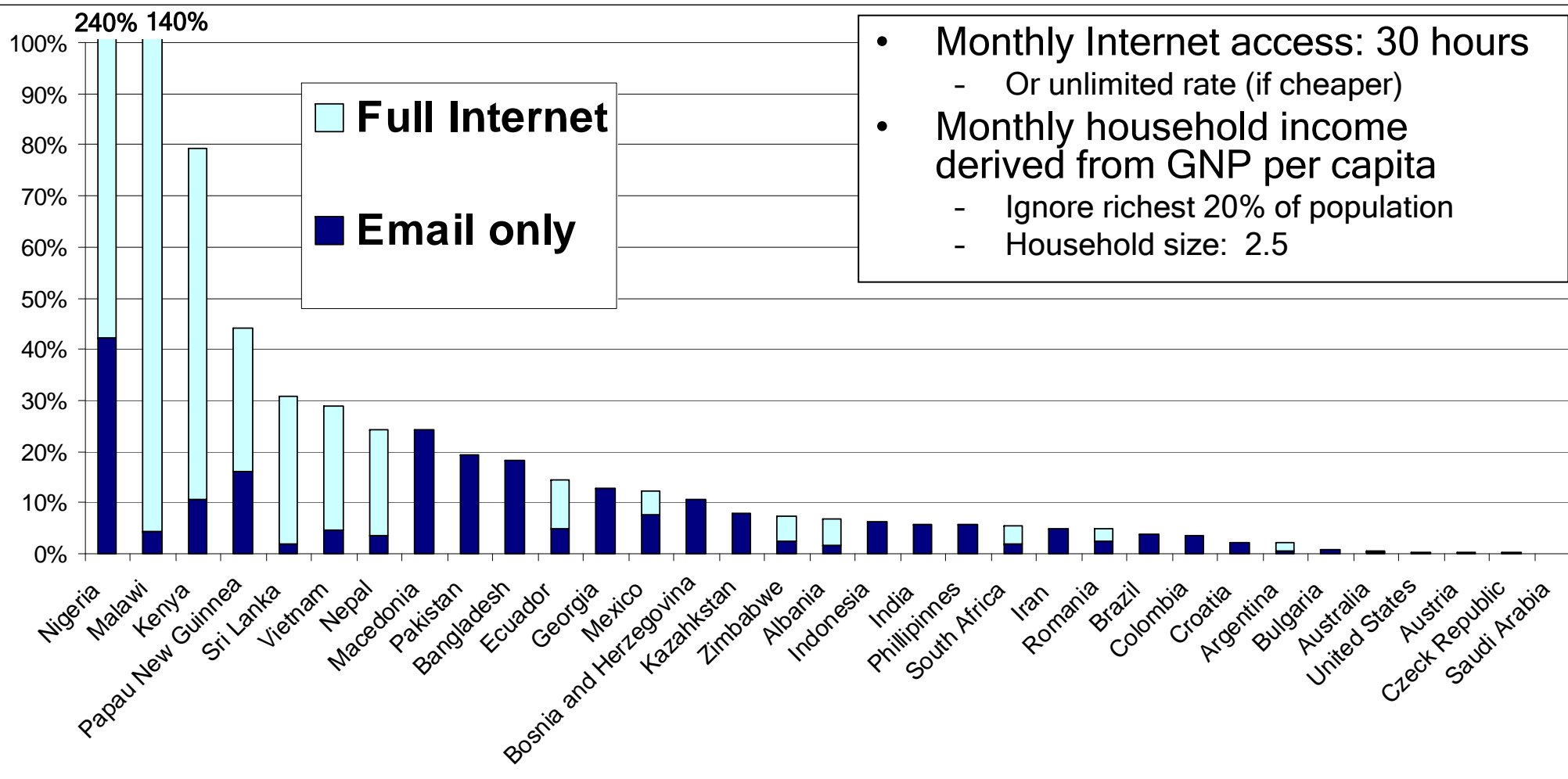
- Infrastructure
  - Limited phone lines
  - Low-bandwidth international links
  - Unreliable power supplies
- High costs
  - Computer unaffordable or unavailable
  - ISP, telephone costs can exceed local wage
  - Exacerbated by slow connections
- Social barriers
  - Illiterate or non-technical users
  - Lack of local content

# Cost of Dial-up Internet Access as a Fraction of Household Income

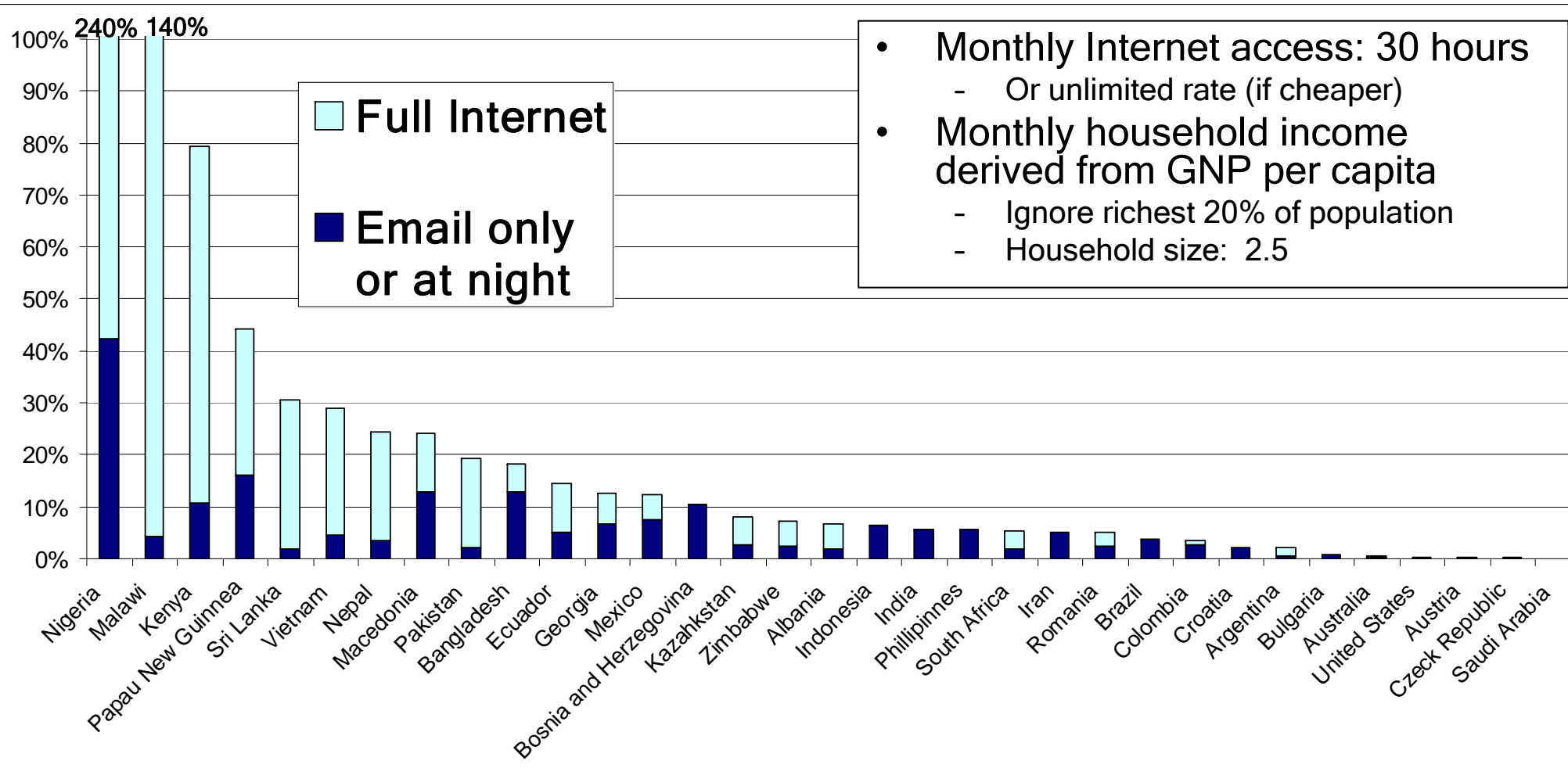




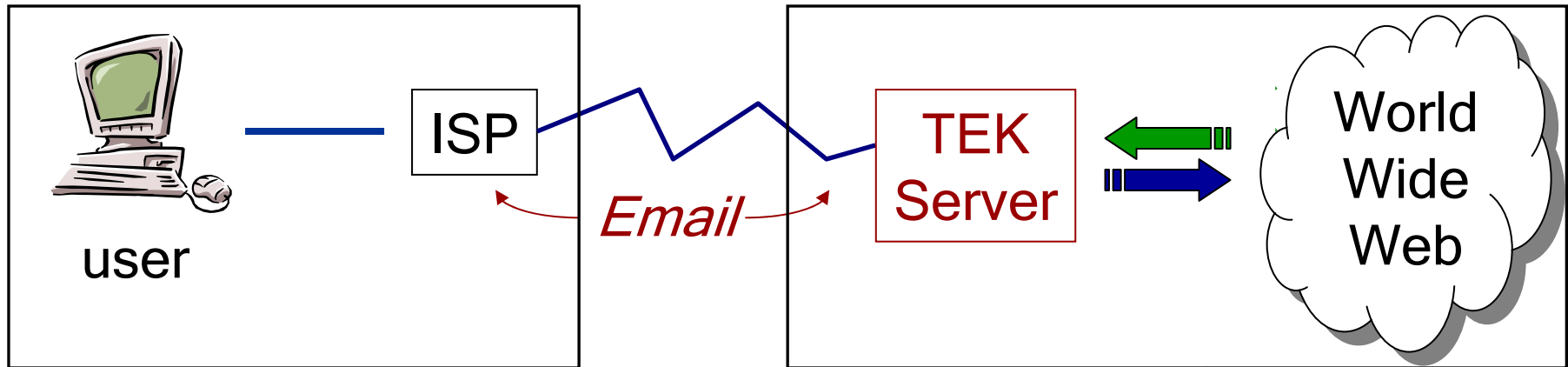
# Cost of Dial-up Internet Access as a Fraction of Household Income



# Cost of Dial-up Internet Access as a Fraction of Household Income



# TEK: Email-Based Search

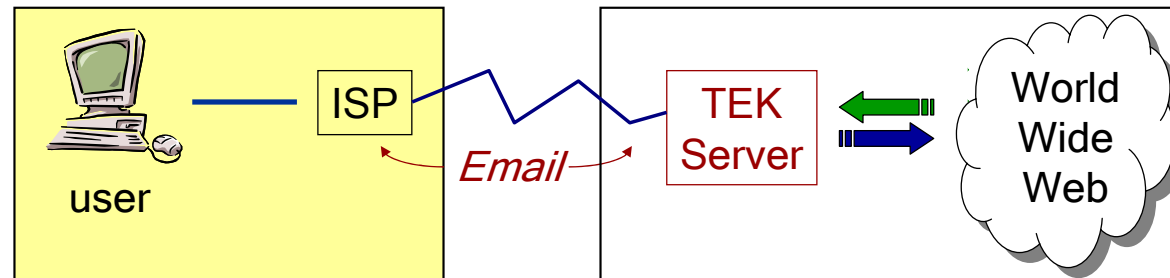


Solution has two components:

1. Transfer all data through email, not http
  - Connect only to send/receive email, not to browse web
2. TEK Server optimizes for bandwidth requirements

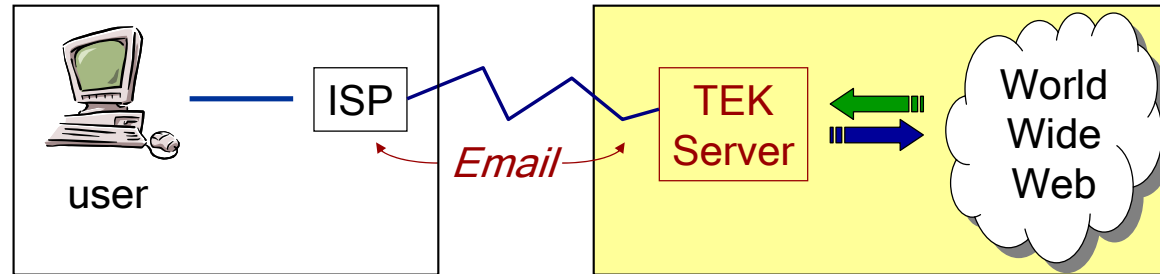
*TEK: "Time Equals Knowledge"*

# TEK Client



- Implemented as an HTTP Proxy Server bundled with a custom version of Firefox
- When offline, users can:
  - Search and browse old results as if connected
  - Enqueue queries for new results or missing pages
- When online, users can:
  - Send pending queries
  - Receive new results (attached to standard emails)

# TEK Server



- Queries Google or Scirus for relevant pages
- Returns filtered content of ~20 pages to user
  - Remove images
  - Remove junk HTML (JavaScript, colors, meta tags, etc.)
  - Convert PDF, PS to HTML
- Compress pages, send as single attachment
  - Overall size reduction: 5-10X on HTML, 10X on PDF
- Maintain server image of client page cache
  - Avoid sending duplicate pages



Go

**TEK**  
TIME EQUALS KNOWLEDGE

Private Account of Bill Thies

[Logout](#)**Search**[Queries](#)[Results](#)[Help](#)

Search for:

Not:

E.g. "heart attack" AND stress

Get this URL:

tsunami aid

[Previous results\(4\)](#)

united nations g8 summit

[Previous results\(30\)](#)© MIT TEK & [Scirus](#) 2005



TEK - Results - Firefox CE TEK

File Edit View Go Bookmarks Tools Help

http://tek/results2.html?term=aids+treatment&not=&main=relevance&page=1 Go

TEK Home

**TEK**  
TIME EQUALS KNOWLEDGE

Private Account of Bill Thies  
[Logout](#)


[Search](#) [Queries](#) [Results](#) [Help](#)

Search for:  Not:  [Local search](#)

Get this URL:  [Local Search](#)

**60 Results**  
Sort by: best results | [newest results](#)

1. [HIV symptoms and information on treatment and s...](#)  
<http://www.managinghiv.com/>
2. [HIV/AIDS Prevention](#)  
<http://www.avert.org/hivprevention.htm>
3. [HIV/AIDS - aids, hiv, aids symptoms, hiv testin...](#)  
<http://bc.us.yahoo.com/b?P=Si2FO9htdWIn0BF962aNv...>
4. [Newly Diagnosed: HIV/AIDS Symptoms](#)  
[http://mv.webmd.com/content/pages/11/1624\\_50945....](http://mv.webmd.com/content/pages/11/1624_50945....)

Find:  Find Next Find Previous Highlight Match case  Phrase not found

Done



# Testing Positive for Aids/HIV

[Your Guide, Tracee Cornforth](#)  
[i-Your Guide, Tracee Cornforth](#)

From [Tracee Cornforth](#),  
Your Guide to  
[Women's Health](#).  
**FREE** Newsletter. [Sign Up Now!](#)

**What does testing positive for HIV mean? What is meant by the window period? How does a false positive relate to it?**

A window period is a recommended waiting period to receive an accurate HIV test result. Generally, it is a six-week to six-month period from the moment of your last unsafe sex encounter to the moment that you receive a HIV screening. This is the time your body uses to create antibodies in the blood stream, which signify exposure to HIV. This process is known as seroconversion.

It is important when receiving an HIV test to ask what kind of test is being used. Whenever someone is screened for HIV, two types of tests are used. They are, 1) a reactive test, and 2) a confirmatory test. A reactive HIV test indicates if HIV antibodies are in the blood (such as the Elisa Test). A reactive test may give a false positive reading to

[About.com](#)[i-About.com](#)

[Health & Fitness](#)

[Women's Health](#)

[Essentials](#)

[What Do My Symptoms Mean?](#)[Women's](#)

[Health Treatments](#)[Frequently Asked Female](#)

[Health Questions](#)[What You Need to Know](#)

[About Your Health](#)[Most Requested Women's](#)

[Health Articles](#)

[Articles & Resources](#)

[Heart Health for Women](#)[Menopause -](#)

[Perimenopause](#)[Sexual Health](#)[Birth](#)

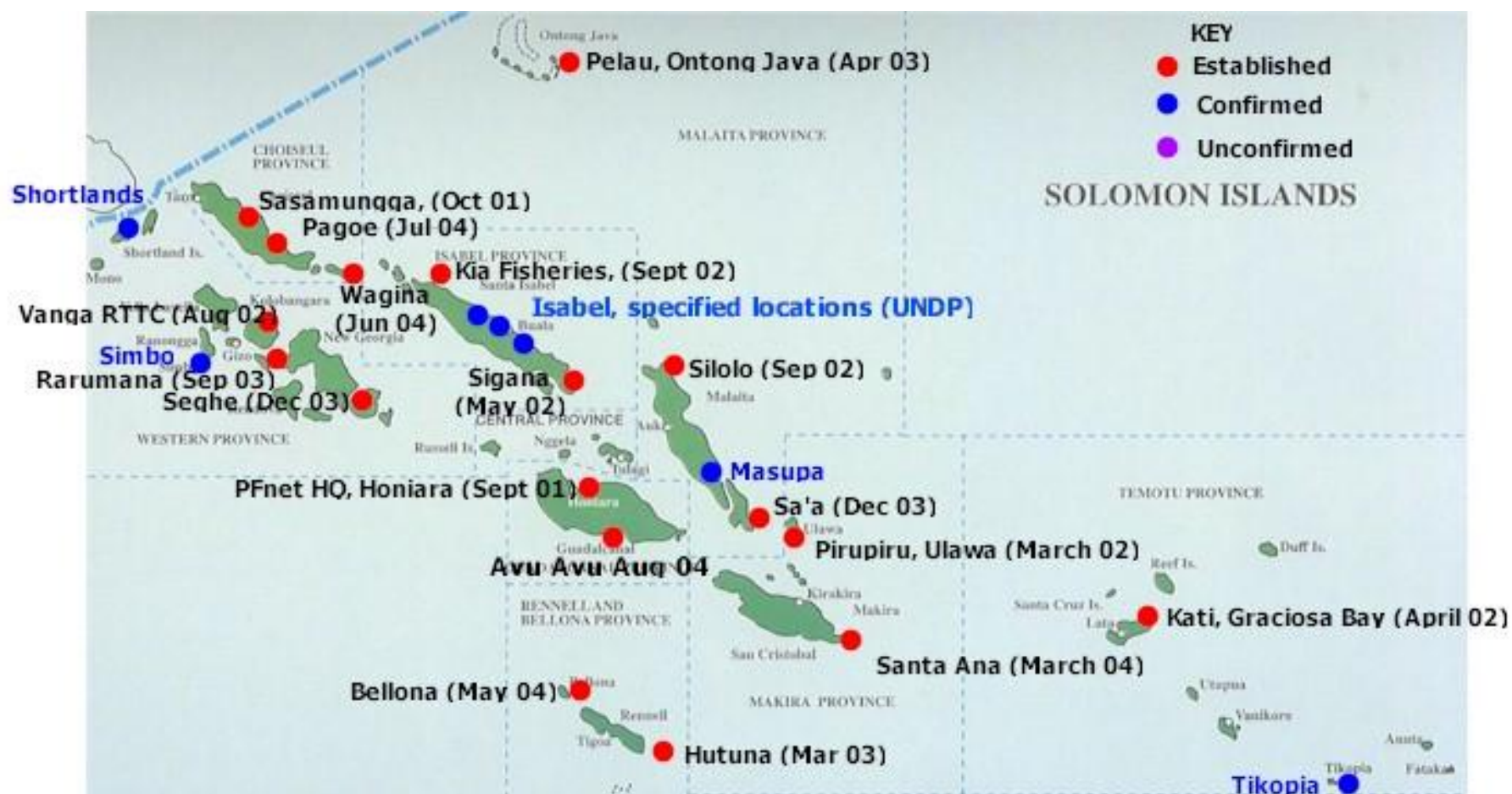
# TEK Users



- TEK available on SourceForge and via free CD
- Most active users in partner organizations
  - People's First Network
  - United Villages

# People's First Network

- Solomon Islands served by HF Radio Network
- Email only



Source: [http://www.peoplefirst.net.sb/General/PFnet\\_Update.htm](http://www.peoplefirst.net.sb/General/PFnet_Update.htm)

# People's First Network

- TEK installed: \$0.65 per query from kiosk
  - Compare to \$0.25 per email, \$0.65 to type one page
  - \$1.30 / hour for operator assistance browsing results
  - **Contributes to kiosk sustainability**
- Many applications reported
  1. Farmers - information on diseases; networking

Subsistence farmers on Rennell have obtained advice concerning taro diseases affecting their crop. Via the 'TEK-websearch' facility, one group of farmers was able to access detailed technical information about vanilla farming and to communicate with a specialist from the *Kastom Gaden Association*. -- Chand et al., *PFNet Case Study, 2005*
  2. Teachers - environmental impact of local logging
  3. Pastors - downloading sermons
  4. Entrepreneurs - download / sell lyrics
  5. General - health, education, sports, entertainment



# United Villages

- Store-and-forward connectivity via Mobile Access Point



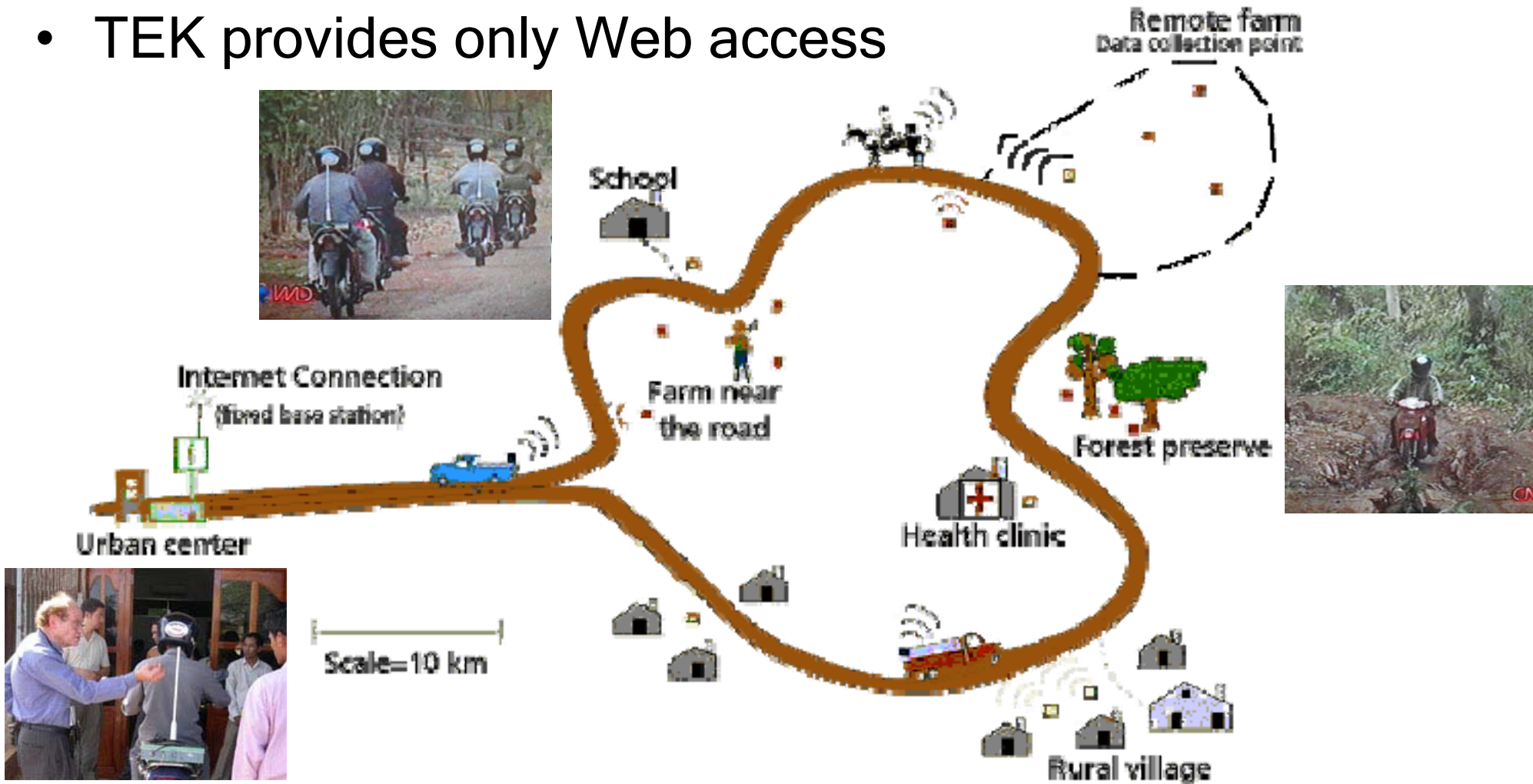
# United Villages

- Store-and-forward connectivity via Mobile Access Point



# United Villages

- Store-and-forward connectivity via Mobile Access Point
  - India, Costa Rica, Cambodia, Rwanda, Paraguay
- TEK provides only Web access



# Applications in Developed Countries

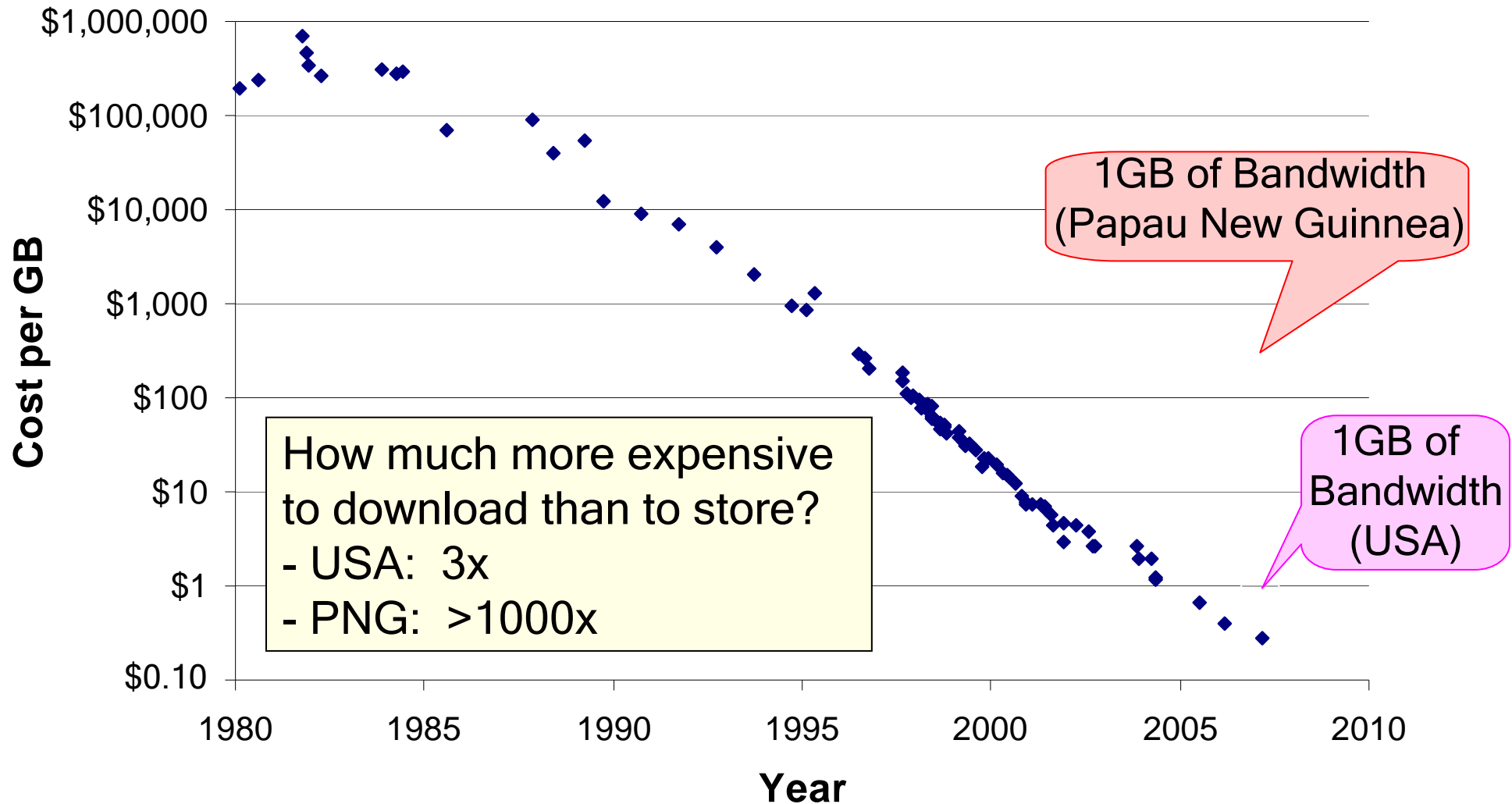
- Airplanes
  - Tenzing supplies email-only connection (\$10-\$20)
    - Continental Airlines, United Airlines, US Airways
  - ~2.4kbs satellite link for entire plane<sup>1</sup>
- Mobile phones
- ISPs charge for bandwidth (Australia)
- Conservative religious sects<sup>2</sup>
- Anxiety about browser security<sup>2</sup>

<sup>1</sup> <http://www.pcworld.com/news/article/0,aid,114216,00.asp>

<sup>2</sup> *anecdotes from EmailWeb users*

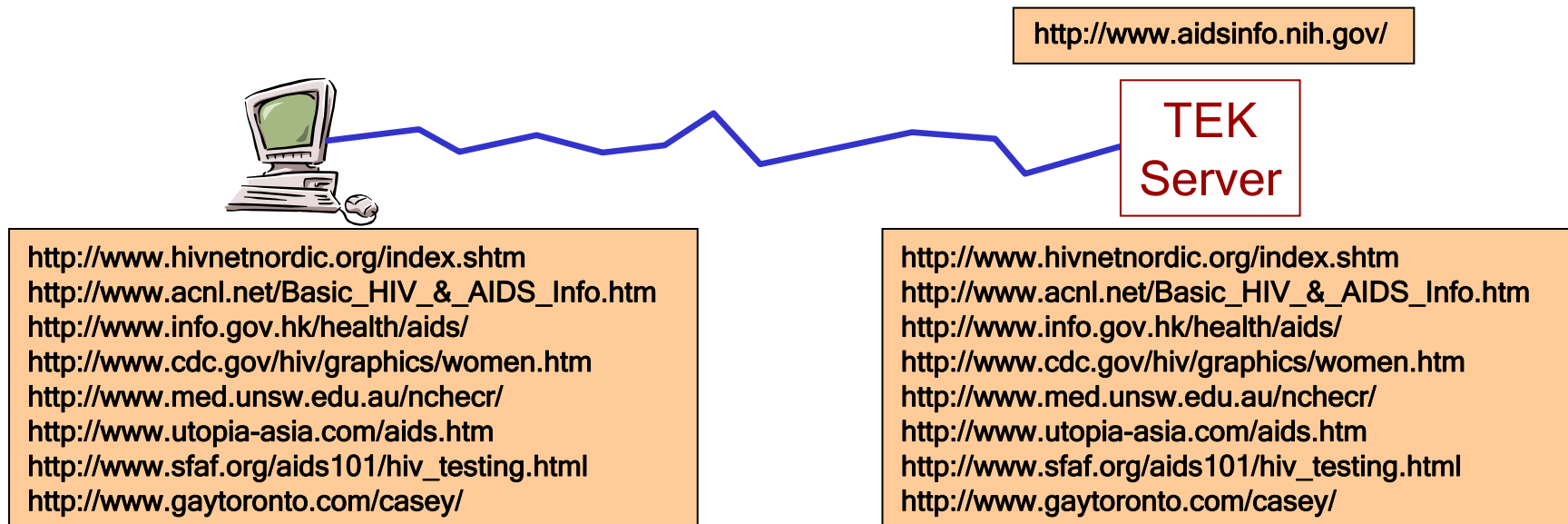


# Cost of Storage



# State-Based Compression

- Idea: use client-side storage reduce bandwidth requirements
- If server knows everything stored on client, can it improve compression of search results?



# State-Based Compression

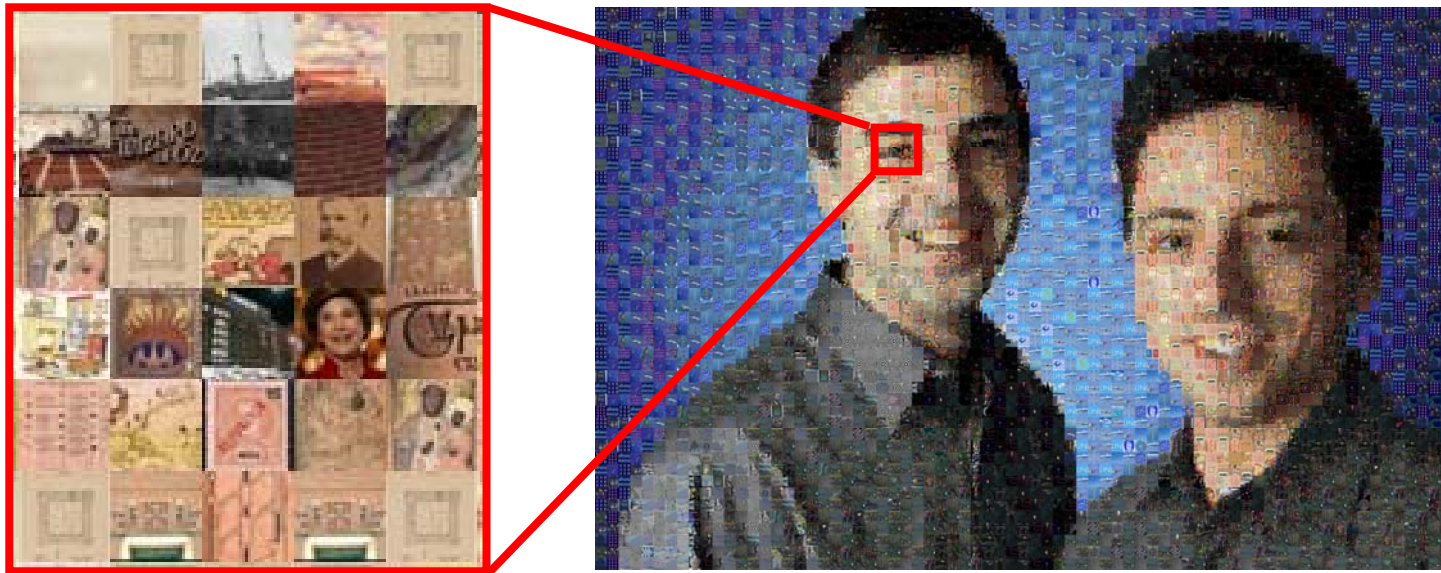
- General problem:  
If two parties share a large dictionary, can they reduce communication bandwidth?



- In general: no
  - info content (index) = info content (entry)
- In practice: maybe
  - Space of inputs is not uniformly populated
    - E.g., many images are text, bullets, smileys, patterns
  - Lossy: send index of closest match in dictionary
  - Lossless: send exact diff from dictionary entry

# Photo Mosaics

- Mosaic: picture made of other pictures

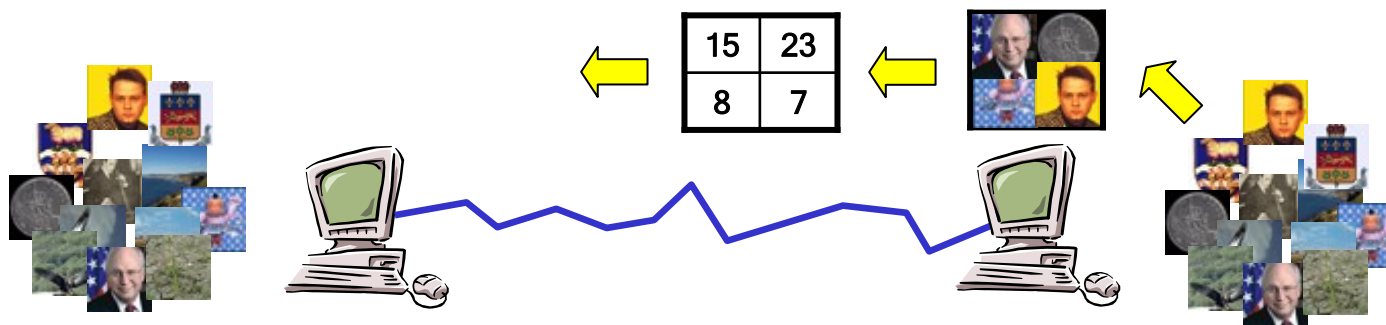


1. Break image into cells
2. Match each cell against image library
  - Use wavelet decomposition for perceptual match

# Mosaic Compression

(Samidh Chakrabarti)

- Idea: server constructs mosaic from client images
  - Send pointers to image components, not image data

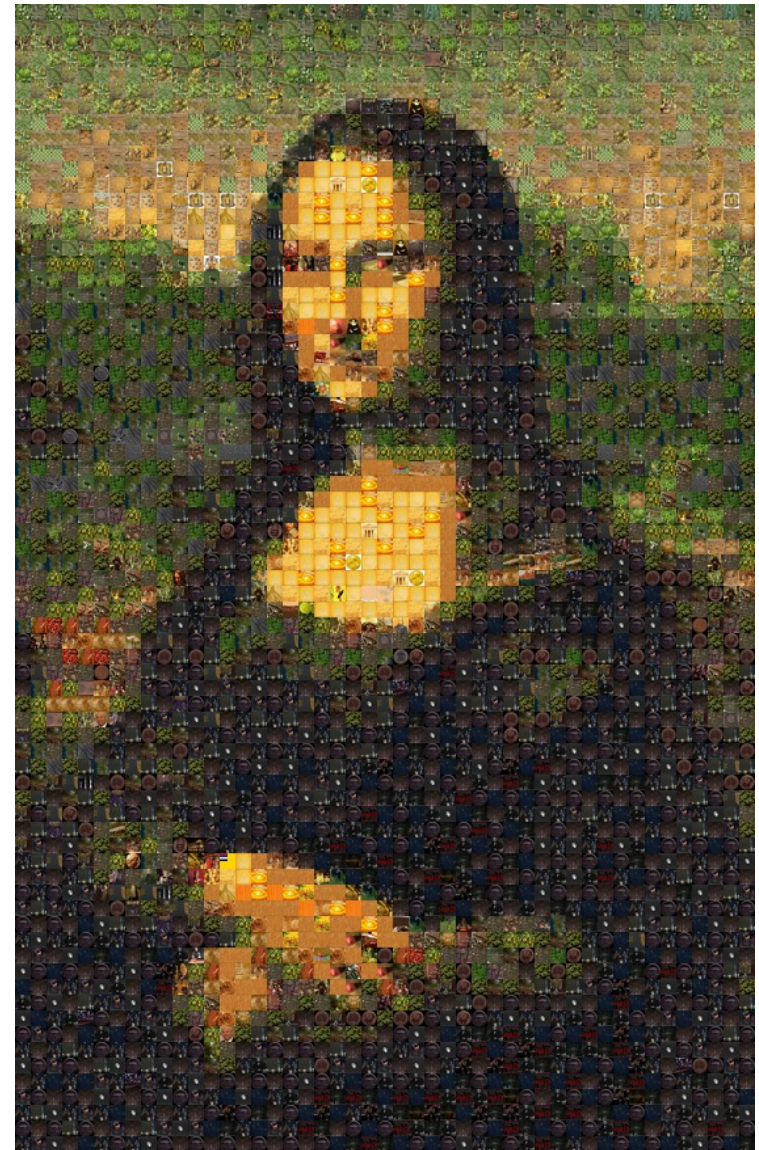


- Experiments
  - Image library: 4096 images from Wikipedia
  - Use shareware PhotoMosaic software (BlackDog)





Wikipedia JPEG: 46 Kb

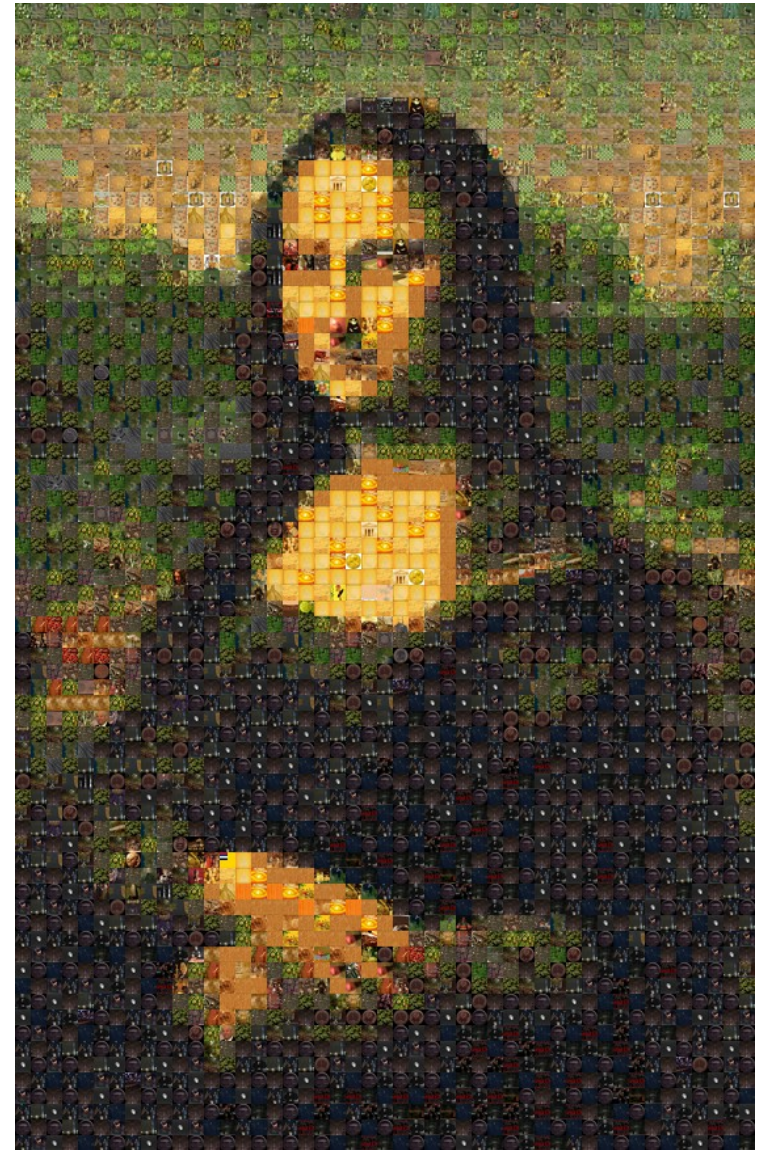


Mosaic: 2.0 Kb  
(22X smaller)





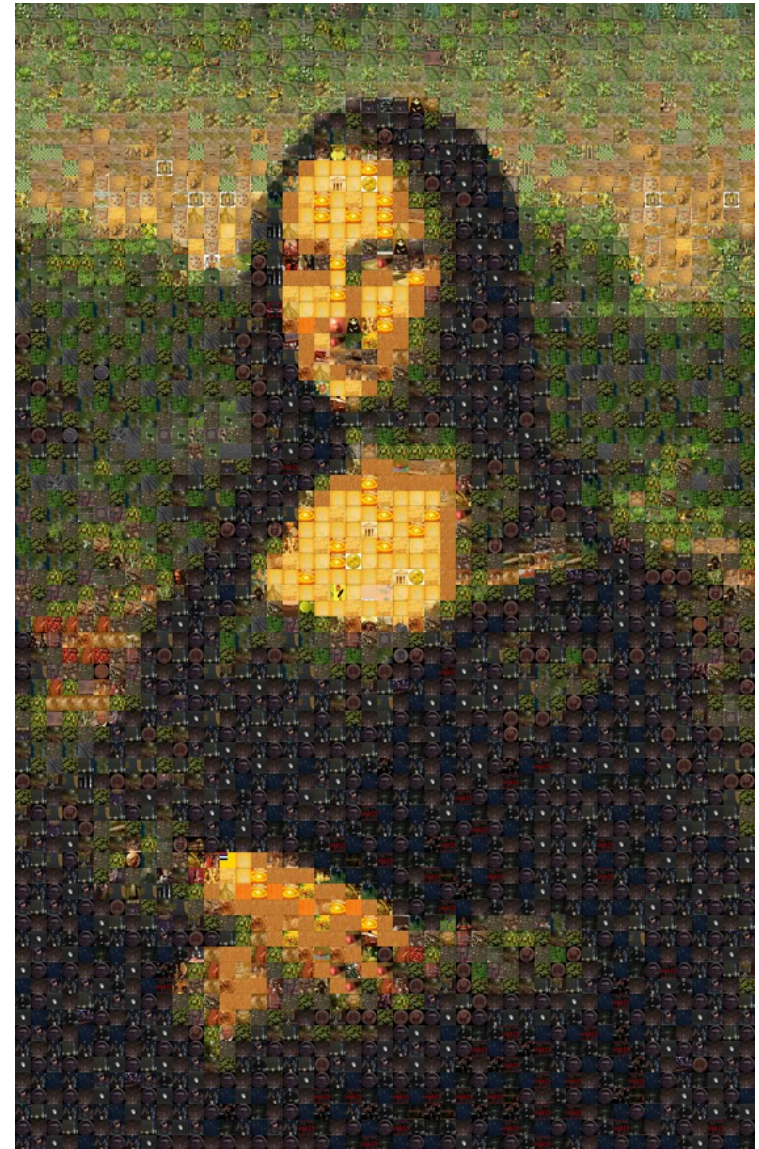
2.0 Kb JPEG



Mosaic: 2.0 Kb



2.0 Kb GIF



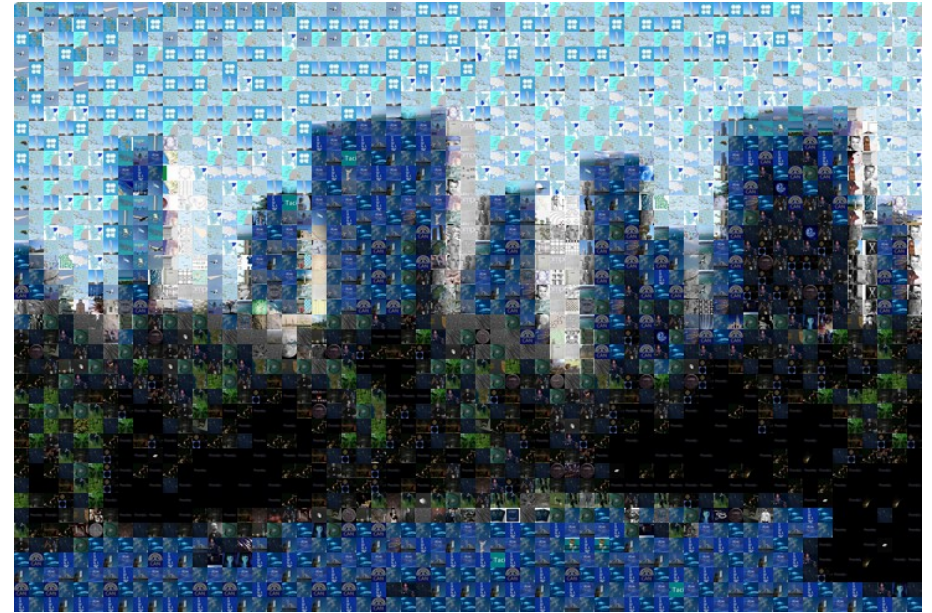
Mosaic: 2.0 Kb



# Compressing Landscapes

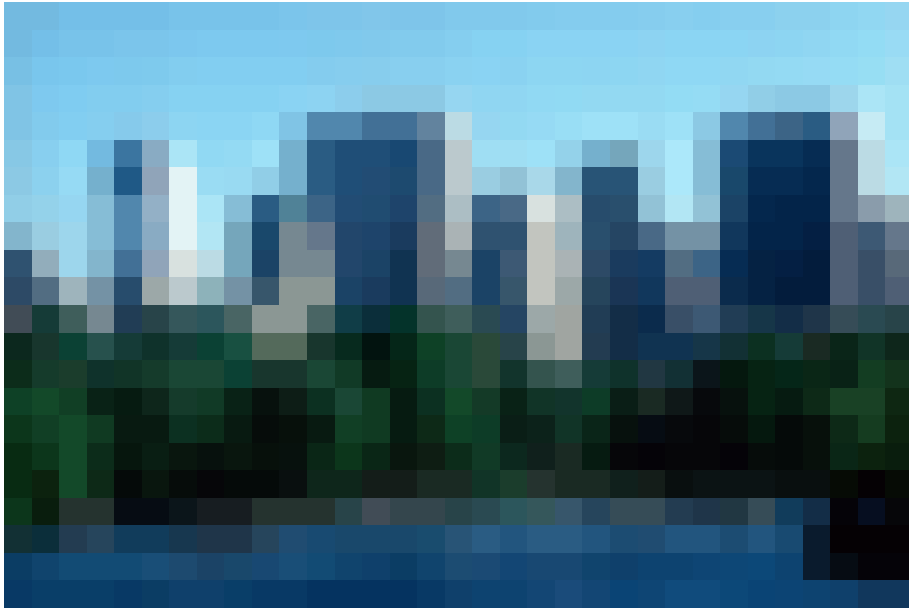


JPEG Image: 52 Kb

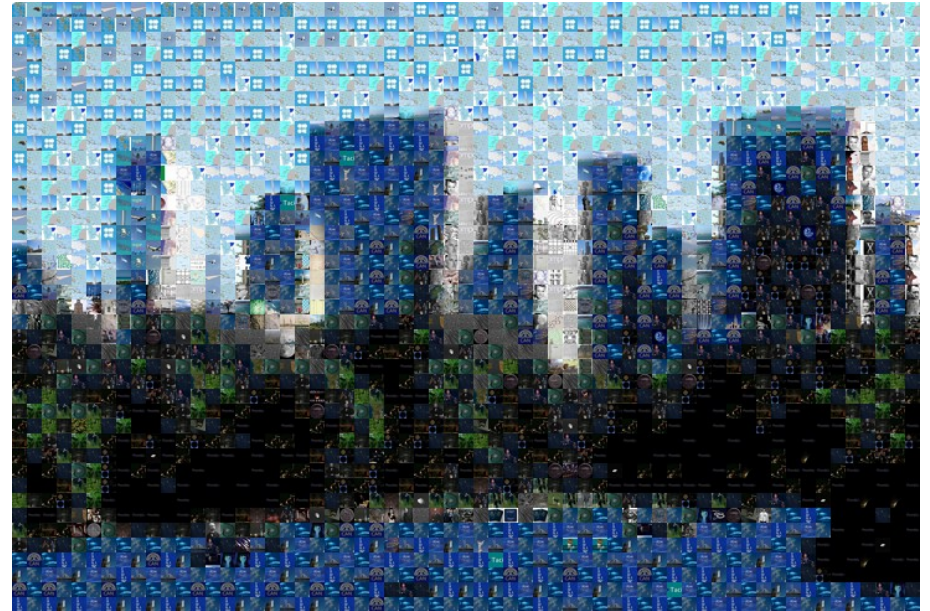


Mosaic: 1.6 Kb  
(33X smaller)

# Compressing Landscapes



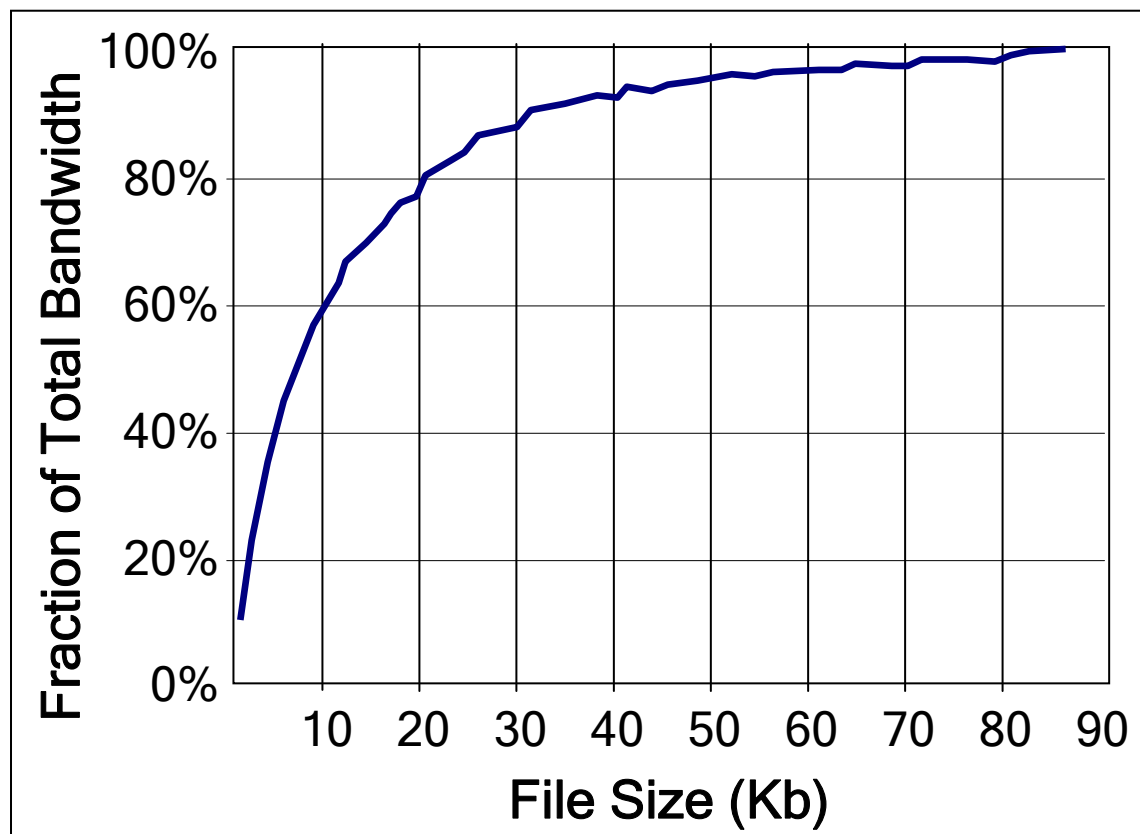
1.6 Kb GIF



Mosaic: 1.6 Kb

# Importance of Small Images

- Most bandwidth spent on small images!



60% of bandwidth  
on images < 10Kb

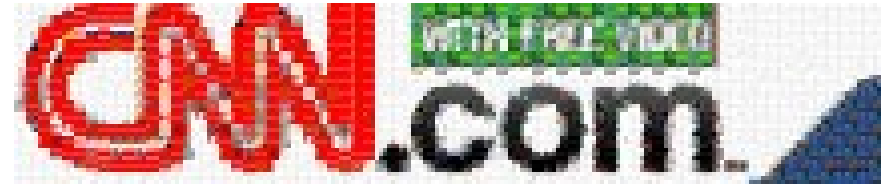
→ Text recognition?  
→ Icon substitution?

- Source: Chakrabarti'02
- 42,684 images from sites in Google programming contest
- 5,540 images from 1,000 most popular sites (ZDNet)

# Compressing Logos

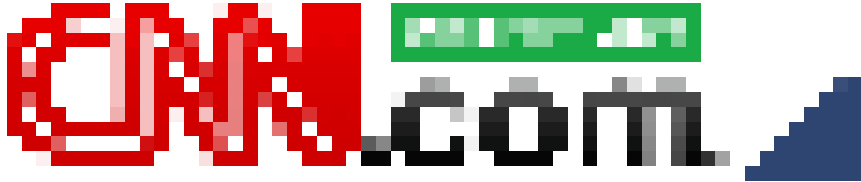


Original GIF: 4 Kb

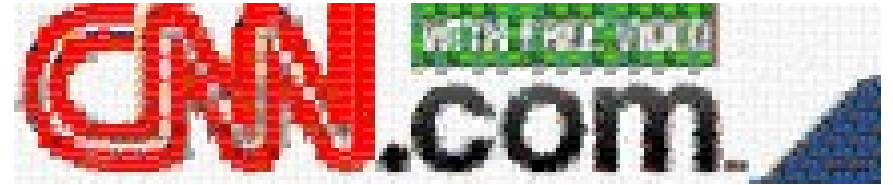


Mosaic: 0.8 Kb  
(5X Smaller)

# Compressing Logos



3.7X Smaller GIF: 0.8Kb



Mosaic: 0.8 Kb  
(5X Smaller)

# What's the Verdict?

- Many avenues for improvement
  - What is the best image library?
  - Impact of smoothing, rotation, diffs?
  - Edge detection + texture mapping
    - Lossy compression of edges
    - Random noise for realism
- In current form, perhaps useful as a preview
  - 5-33X smaller than JPEG
  - More entertaining than ALT tag or blurry picture



CNN.com



# In This Talk

1. Retrieving information
- 2. Publishing information**
3. Collecting medical data



# The Power of Publishing

- **User-Generated Content has come to define the Web**

- Original attraction of the Web....everyone can be a publisher
- Now...Blogs, review sites, digital video, forums, news comments, ...
- Empowers ordinary citizens with a voice + a global audience

**“75% of all content on the Web is user-generated.”**

— Reggie Bradford, CEO of Vitruve

**“35% of U.S. Internet users have posted some sort of user-generated content online.”**

— Home Broadband Adoption 2006, Pew Internet & American Life Project

- **However, the ability to publish and share content remains out-of-reach for most people around the world**

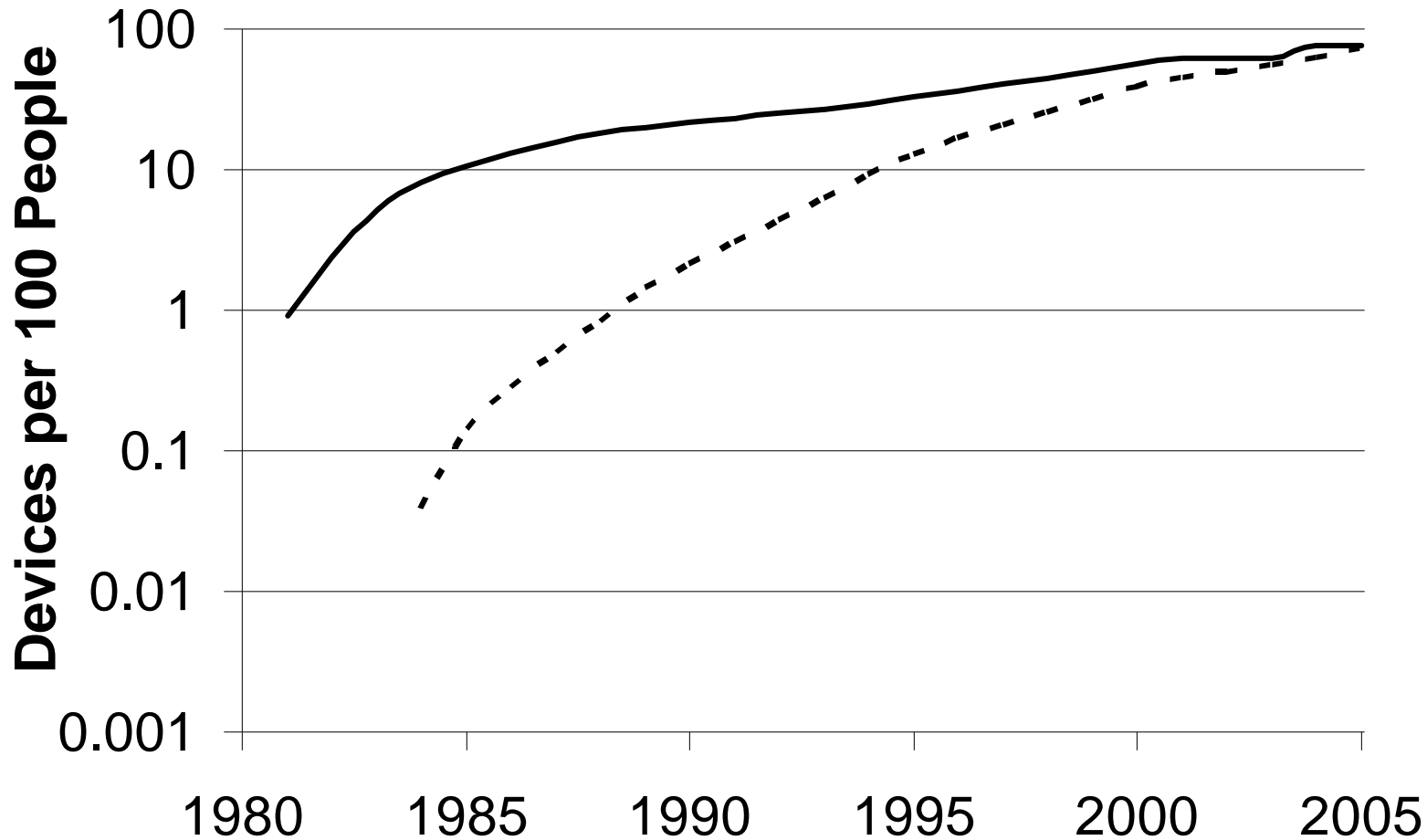
1. Lack of personal computers, Internet access
2. Language/literacy barriers

→ Can we use cell phones + voice to enable content sharing?



# 1. Overcoming the Lack of Computers

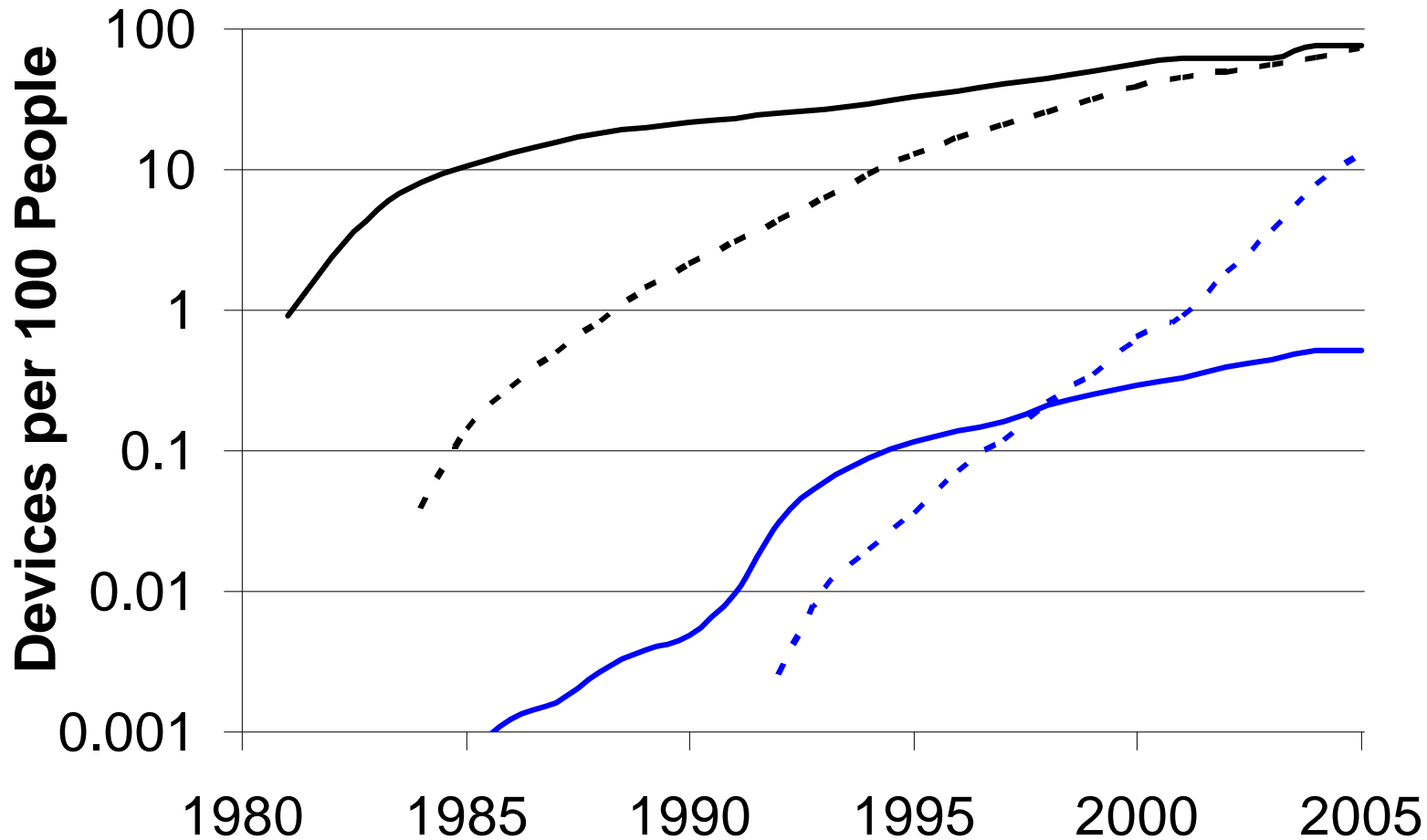
— PCs (USA)  
- - - Cell phones (USA)



*Source: International Telecommunications Union*

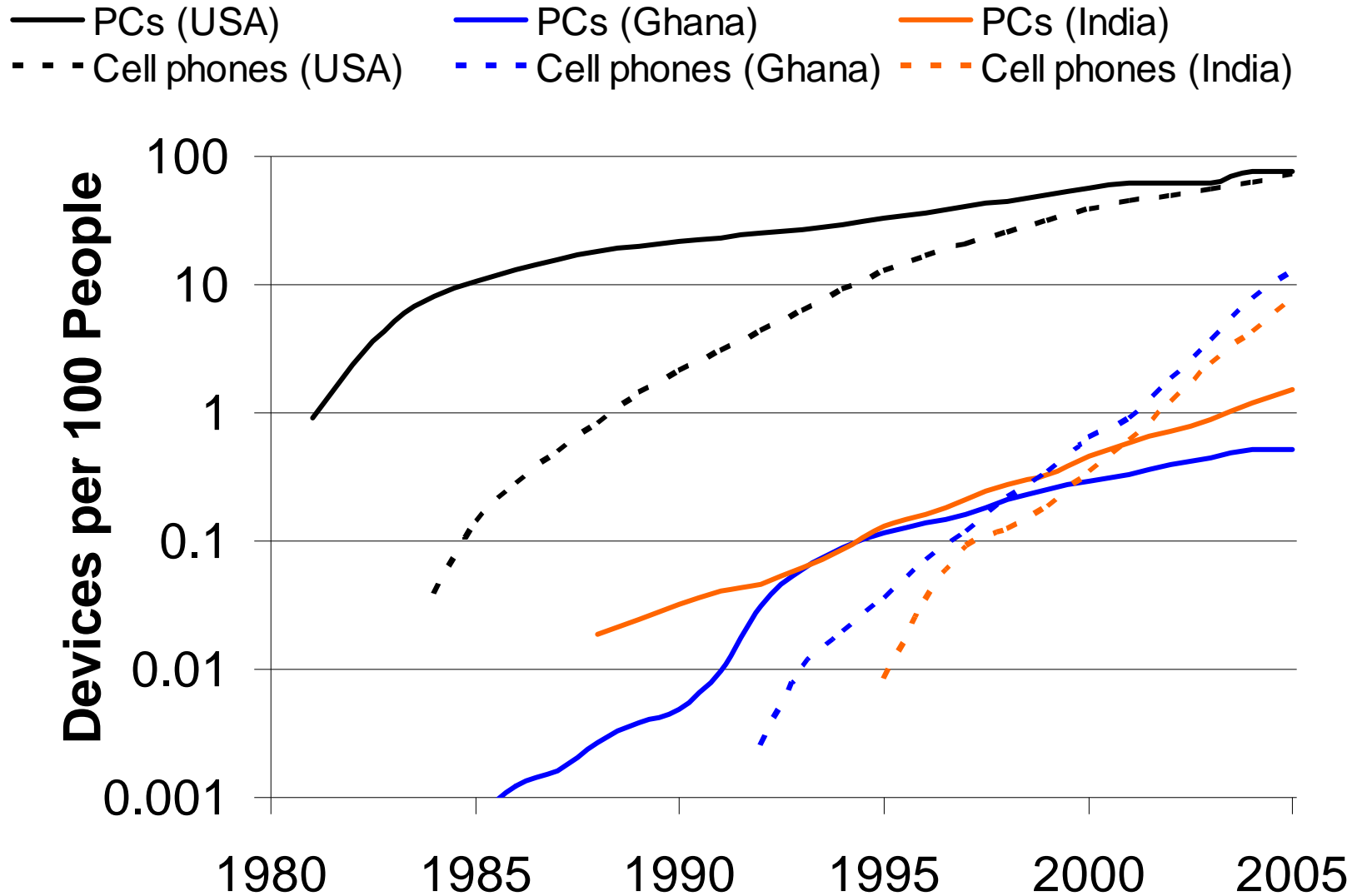
# 1. Overcoming the Lack of Computers

— PCs (USA)      — PCs (Ghana)  
- - - Cell phones (USA)      - - - Cell phones (Ghana)



*Source: International Telecommunications Union*

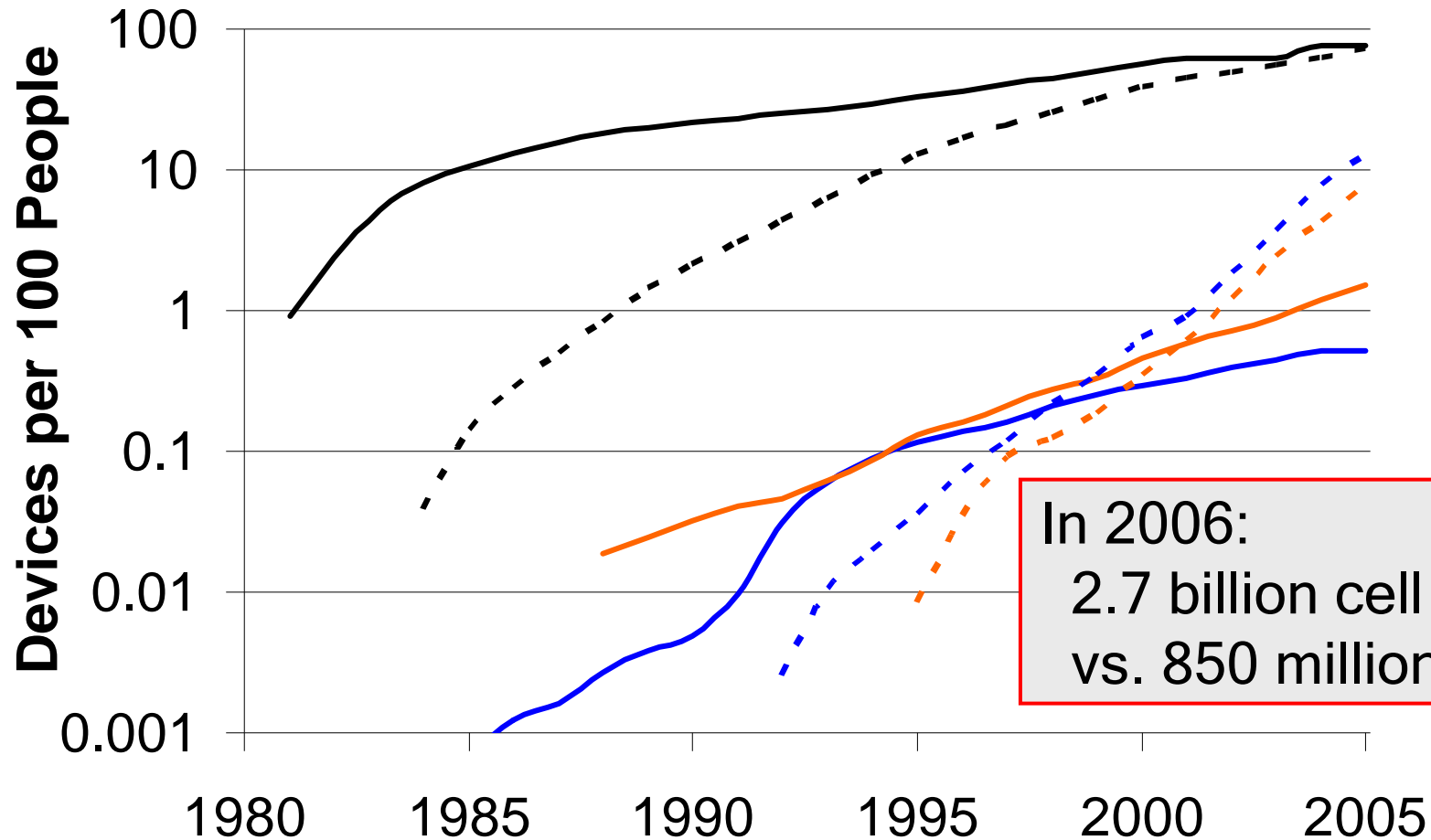
# 1. Overcoming the Lack of Computers



Source: International Telecommunications Union

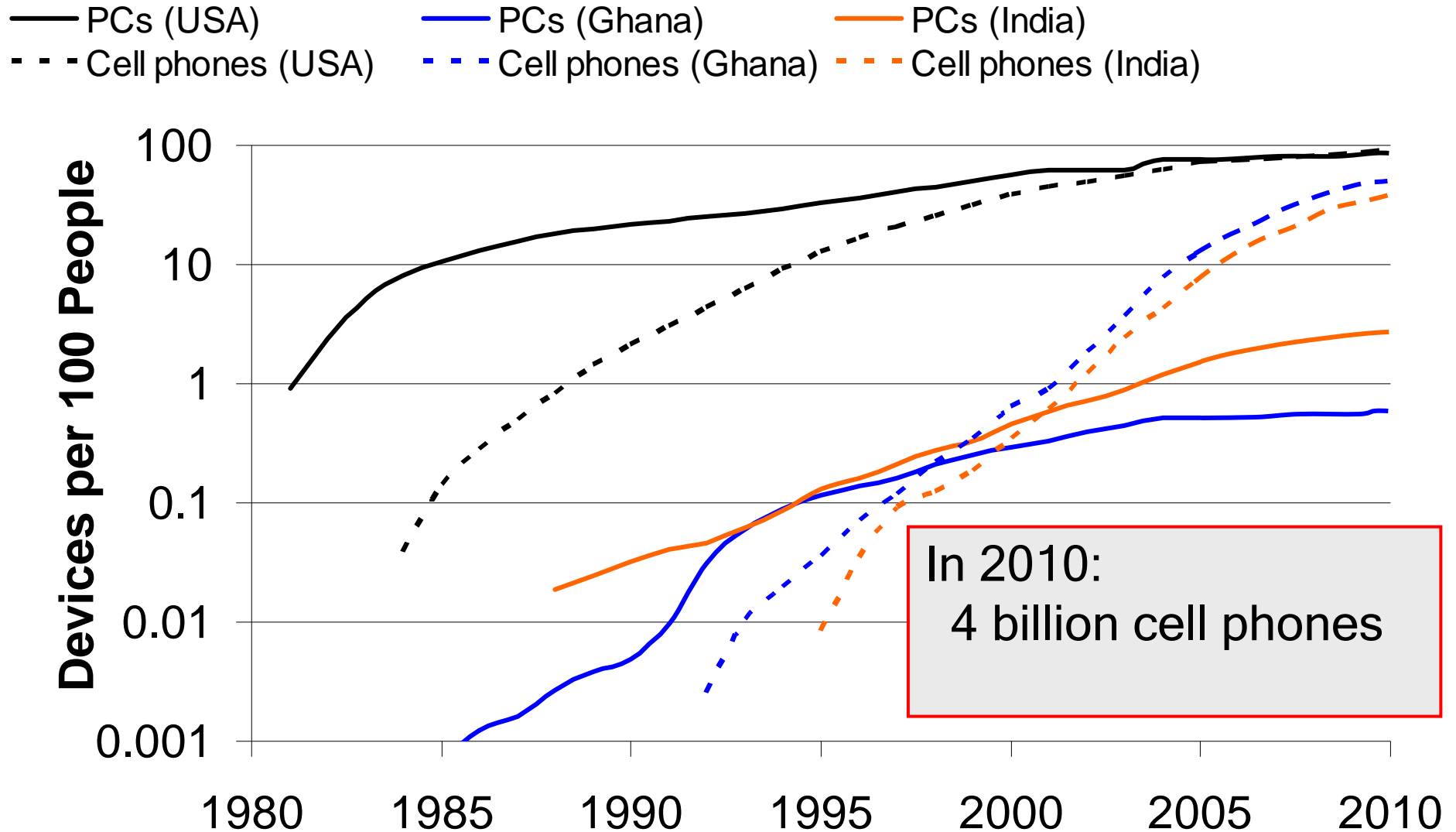
# 1. Overcoming the Lack of Computers

— PCs (USA)      — PCs (Ghana)      — PCs (India)  
- - - Cell phones (USA)      - - - Cell phones (Ghana)      - - - Cell phones (India)



In 2006:  
2.7 billion cell phones  
vs. 850 million PCs

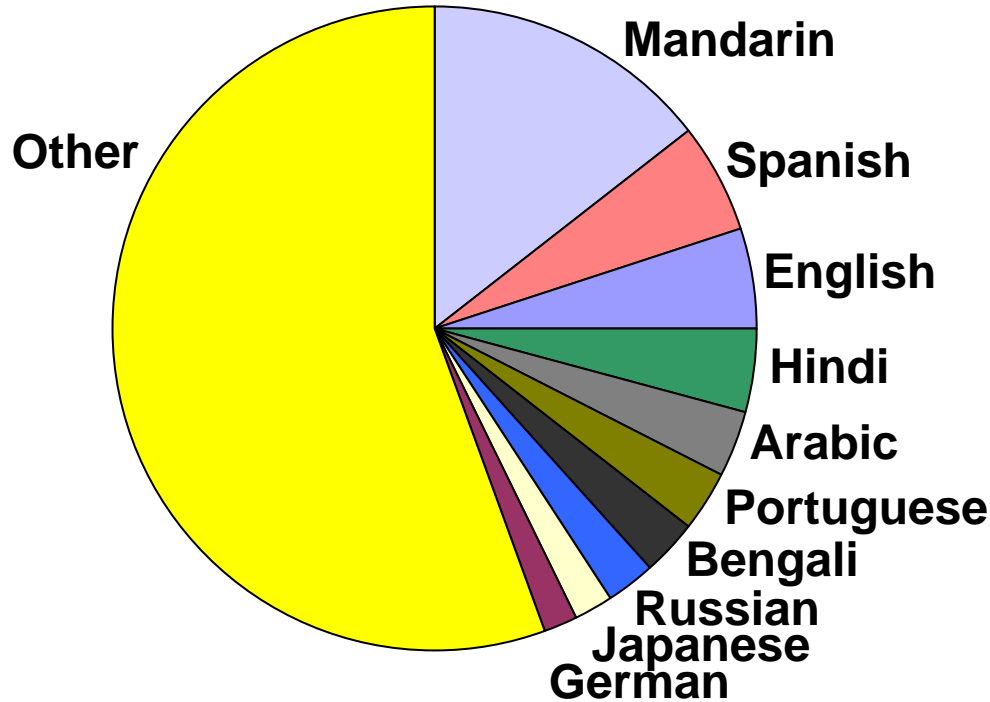
# 1. Overcoming the Lack of Computers



Source: International Telecommunications Union

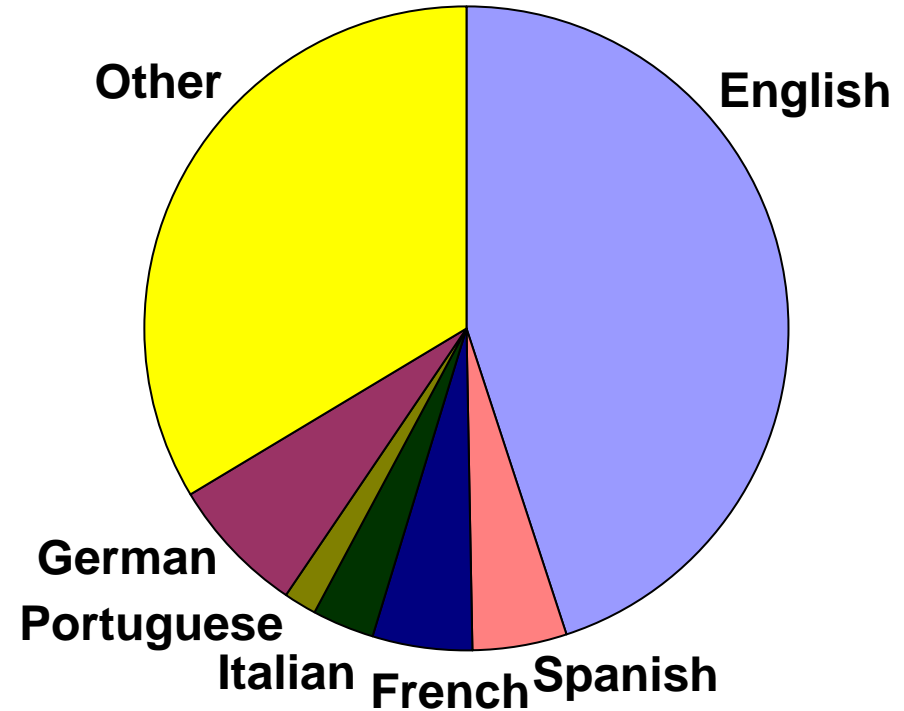
## 2. Overcoming Language Barriers

- **Top languages:**



***By native speakers***

*(Source: Wikipedia, 2008)*



***By Internet Web Pages***

*(Source: Networks & Development Foundation, 2005)*

- **Existing interfaces heighten language barriers:**

- Text displays require literacy
- Text input requires complex font encoding

# Solution: An Audio Wiki

- **Allow users to publish information:**
  - Using a phone rather than a computer
  - Using voice rather than text
- **Audio recording and playback, but keypad-driven navigation**
  - Not attempting a dialogue-based system
- **Research challenge: making it usable**
  - Interactive voice response (IVR) typically frustrating
  - Research: adaptive interfaces, audio linking, flexible playback
- **Two prototypes in development:**
  - Structured forum on a given topic (ala Yahoo! Answers)
  - Completely flexible Wiki (ala Wikipedia)



*recording,  
playback*



*navigation*

# Prototype 1: Structured Forums

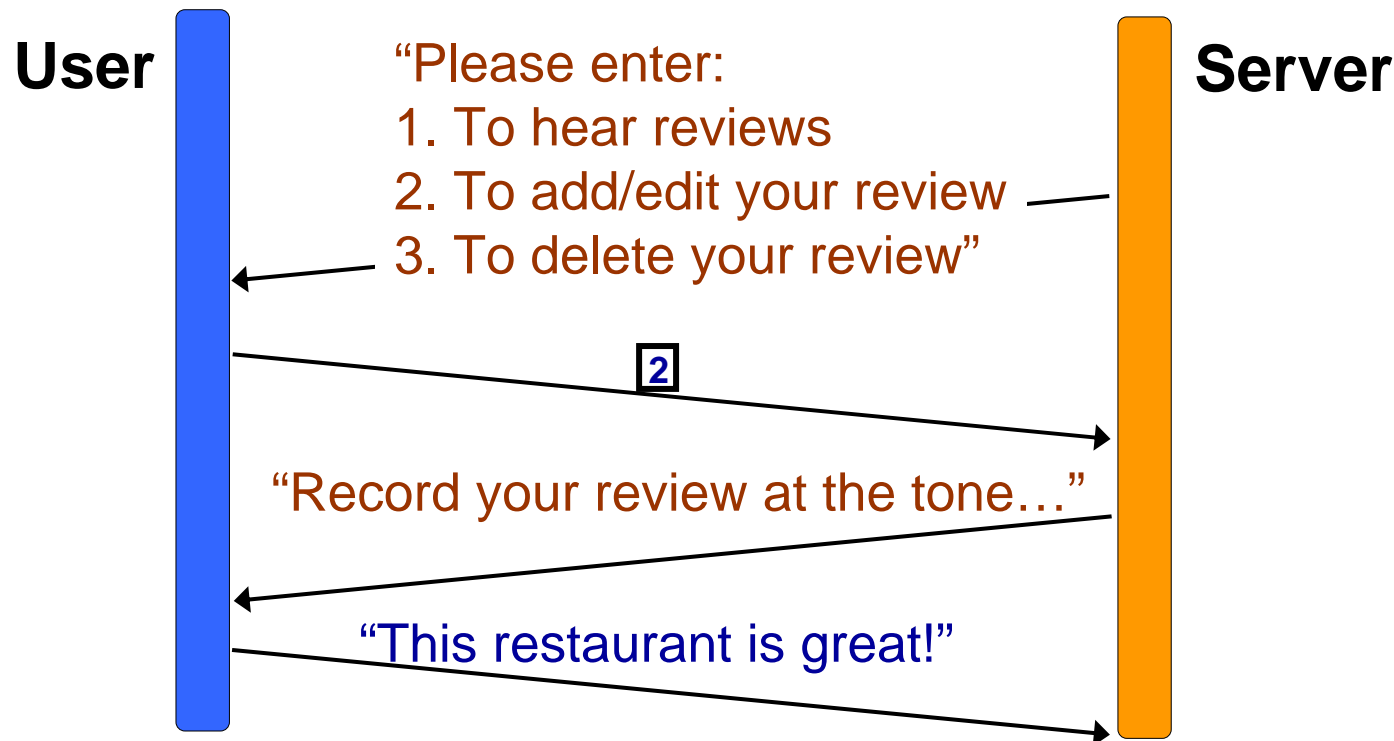
- **Organized as set of topics with user comments**
  - Users can:
    - Add topics, search for topics (using keypad)
    - Add, edit, or listen to comments (using keypad+voice)
- **Initial testbed: restaurant reviews**





# Prototype 1: Structured Forums

- **Organized as set of topics with user comments**
  - Users can:
    - Add topics, search for topics (using keypad)
    - Add, edit, or listen to comments (using keypad+voice)
- **Initial testbed: restaurant reviews**



# Ranking and Improving Content

- **User-generated content sites rely on ratings and incentives to improve the quality of content**
  - Article of the month (Wikipedia)
  - Featured reviewers (Amazon, Yelp, etc.)
  - Highest-rated contributors (YouTube, Photo.net, etc.)
  - Financial compensation (Metacafe, Revver, etc.)
- **Mechanisms for evaluation / ranking on an audio wiki:**
  - 1. *Track most popular content***
    - Users can press key to skip remainder of entry
    - Least-skipped entries are played first
    - Such tracking is more difficult to perform on the Web
  - 2. *Leverage user profiles***
    - Users may create profiles, send/receive voicemail
    - All content is linked to author's profile with press of key
    - Provides incentive to create good online persona

# Prototype 2: Flexible Wiki

- **Unlike the structured forums, allow all users to edit the same audio snippets**
  - Insert, edit, delete, and link
  - Create and delete audio pages
- **How to lookup content?**
  - Spell with keypad
  - Ask 20 questions
  - Language-agnostic search of keywords
- **New user interface abstraction: *the audio cursor***
  - During editing, can advance cursor to point of interest
  - Iterative playback ensures cursor is at intended position
  - Cursor used to mark insertion point, and to select text for deletion or linking

# Implementation Status

- **Prototype implementations complete**
  - Implemented using Asterisk, Summer 2007 – Spring 2008
- **Three deployments underway:**
  - IIT Guwahati: course reviews
  - Boston: restaurant and product reviews
  - Boston: flexible Wiki service
- **Using experience gained, will transition to developing regions**
  - Agriculture
  - Government
  - Entrepreneurship
  - Health

# Agenda for Future Research

- **New interfaces for linking content**
  - Background tone, accented speech, separate section?
  - Need to adapt audio links to unique constraints of Audio Wiki:
    - Limited fidelity
    - Need to edit content
    - Unpredictable user voices
- **Adapting the interface to suite the user**
  - Recognize caller, keep track of preferences, favorites, etc.
  - Allow users to create shortcuts across the user interface
- **Multimodal interfaces**
  - Voice + SMS + local app
  - Voice + web
- **Language independence**
  - Index and search with universal phonemes
- **Improving playback quality**
  - Speed / volume adjustment
  - Noise elimination / equalization
- **Universalization of local content**
  - Find correlations between distant geographies
  - “Content Networking”

# In This Talk

1. Retrieving information
2. Publishing information
3. Collecting medical data



# Innovators In Health: Improving Rural Healthcare Delivery via Data-Driven Intervention

- Tuberculosis remains major challenge in India
  - 4.5 million existing cases
  - 1.7 million new cases per year
  - 300,000 deaths per year

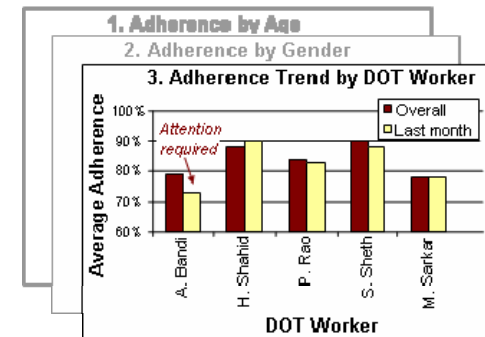
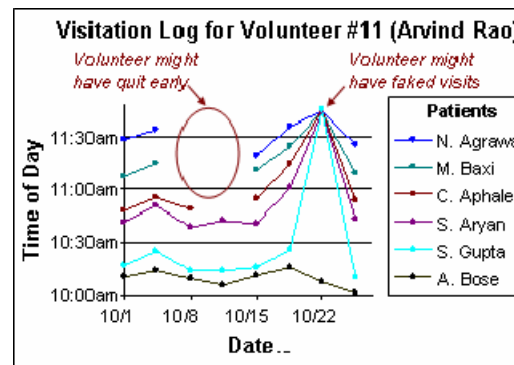
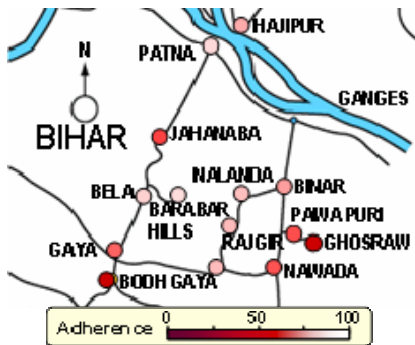


- Logistically challenging to run treatment programs
  - Prajnopaya Foundation's program starting May 2008:  
1600 patients, 45 villages, 80 health workers

➡ **How can we ensure that pills are reaching the patients, and that patients are taking their medication?**

# uBox: A Low-Cost Electronic Pillbox

- Features:
  - Records when pills are dispensed
  - Reminds patient to take pills
  - Locks to prevent overdose
- Benefits:
  - Records of medication adherence can be analyzed in depth

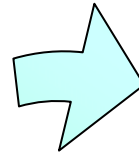


- Enables new incentives, interventions

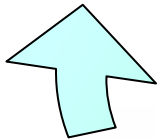
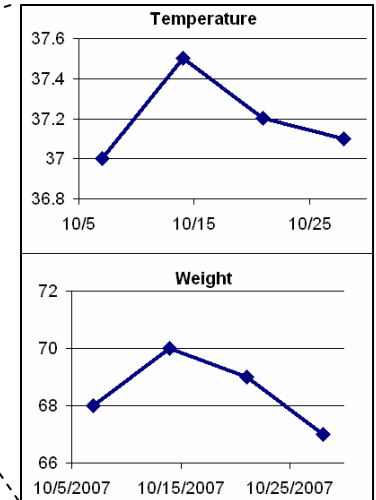
# uPhone: Using Cell Phones to Monitor and Improve Patient Health



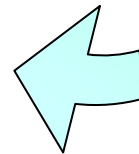
**Health worker**  
enters vital patient health  
indicators into cell phone



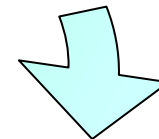
**Nurse**  
analyzes data,  
identifies problems



**Patient**  
lives in a remote area



**Physician**  
sends advice to patients,  
schedules field visits



# ICTD in the MIT Curriculum

- **Fall 2007: Reading group on ICTD (co-organized with Emma Brunskill)**
  - **Readings:**
    - Brewer et al., *The Case for Technology for Developing Regions*
    - Bailur et al., *Review of Research on Rural PC Kiosks*
    - Banerjee & Duflo, *The Economic Lives of the Poor*
    - Jensen, *The Digital Divide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector*
    - Kam et al., *Localized Iterative Design for Language Learning in Underdeveloped Regions: The PACE Framework*
    - Kumar & Best, *Impact and Sustainability of E-Government Services in Developing Countries: Lessons Learned from Tamil Nadu, India*
    - Pawar et al., *Multiple Mice for Retention Tasks in Disadvantaged Schools*
  - **Speakers:**
    - Rich Fletcher (United Villages)
    - Hamish Fraser (Partners In Health)
    - Jonathan Jackson (Dimagi, Inc.)
    - Daniel Kayiwa (Makerere University, Uganda)
    - Kieran Sharpey-Schafer (Cell-Life, South Africa)
- **Spring 2008: 9-unit design class on ICTD**
  - Instructors: Gari Clifford, Rich Fletcher, Jhonatan Rotberg, Luis Sarmenta
  - Cross-listed in EECS, Media Arts & Sciences, Health Sciences Technology

# MIT ICT4D Projects (Spring '08)

(Instructors: Gari Clifford, Rich Fletcher, Jhonatan Rotberg, Luis Sarmenta)

- **Economic Empowerment**
  - Mosoko Mobile Marketplace (Nokia, Kenya)
  - Assured Labor (Brazil)
  - Smart MicroLoans (India)
- **Health**
  - Pediatric care for Urban Poor (Johns Hopkins & IRD, Pakistan)
  - Cervical Cancer Prevention (Dimagi, Zambia)
- **Education**
  - Knowledge Box “Virtual Internet” (Beehive School, Malawi)
- **Community Action**
  - Disaster Management (CRS India)
  - FreePress (Hanantek, Bolivia)



# Related Work

- Delay-tolerant networking
  - TierStore (Demmer, Brewer et al.)
  - ZebraNet (Martonosi et al.)
  - Postmanet, Digital Study Hall (Wang et al.)
  - SeNDT (Geraghty et al.)
  - EmailWeb (Griswold)
  - Interplanetary Internet
- Extreme compression
  - Space-optimized texture maps (Balmelli et al.)
- Audio wiki
  - HealthLine (Sherwani et al.)
  - MobileED (Leionen et al.)
  - World Wide Telecom Web (Kumar et al.)
- uBox / uPhone
  - Pillboxes: e-Pill Dispenser, MedReady Plus, SimPill, AARDEX MEMS, ...
  - Phone programs: Voxiva HealthNet, Cell-Life, Jiva Teledoc, Pesinet, DataDyne EpiSurveyor, MIT MDSS



# Conclusions

- Technology research for the developing world:
  1. Identify unique trends and opportunities
  2. Address with new and appropriate technology

Trend / Opportunity	Research Direction
Email cheaper than Web access	Email-based search
Storage cheaper than bandwidth	State-based compression
Cell phone penetration + Illiterate users	Audio wiki

- Growing community of researchers and educators at MIT

# Acknowledgements

- **TEK**

- Prof. Saman Amarasinghe (PI)
- Libby Levison (PI)
- Tazeen Mahtab
- Genevieve Cuevas
- Saad Shakhshir
- Janelle Prevost
- Mark Halsey
- Marjorie Cheng
- Hongfei Tian
- Damon Berry

- **Mosaic Compression**

- Prof. Saman Amarasinghe
- Samidh Chakrabarti

- **Audio Wiki**

- Prof. Saman Amarasinghe
- Jonathan Birnbaum
- Pratik Kotkar

- **uBox / uPhone**

- Manish Bhardwaj
- Goutam Reddy
- Sara Cinnamon
- Ven. Tenzin Priyadarshi
- Emma Brunskill
- Pallavi Kaushik
- Jessica Leon
- Oliver Venn

