

Knowledge Sharing System

武田英明 Hideaki Takeda

takeda@nii.ac.jp

National Institute of Informatics

Outline of the lecture

- Scope: To learn information and knowledge sharing techniques, such as Knowledge Representation, Semantic Web and Social Network Science
- Category: Artificial Intelligence, Web Informatics
- Learning goal:
 - Conceptual **understanding** of recent development of the knowledge sharing **technologies**
 - Theoretical **understanding** of basic knowledge representation **technologies**
 - Improvement of modeling and programming **skills** to use the knowledge representation and sharing **technologies**
- Prerequisite:
 - Some experience of programming (python, ruby, perl, etc)

Outline of this course

- Topics:
 1. Introduction
 2. Information Integration
 3. Semantic Web: Introduction
 4. Knowledge Representation in AI: Ontology
 5. Knowledge Representation in AI: Description Logics and OWL
 6. Semantic Web Languages: RDF, RDFS, SPARQL
 7. Semantic Web Languages: RDF, RDFS, SPARQL (cont.)
 8. Semantic Web Languages: RDF, RDFS, SPARQL (cont.)
 9. Linked Open Data
 10. Linked Open Data (cont.)
 11. Knowledge Graph
 12. Knowledge Graph (cont.)
 13. Practice
- Style of the lecture
 - 2-3 talks then some small Assignments (exercise and/or presentation)
 - Final report

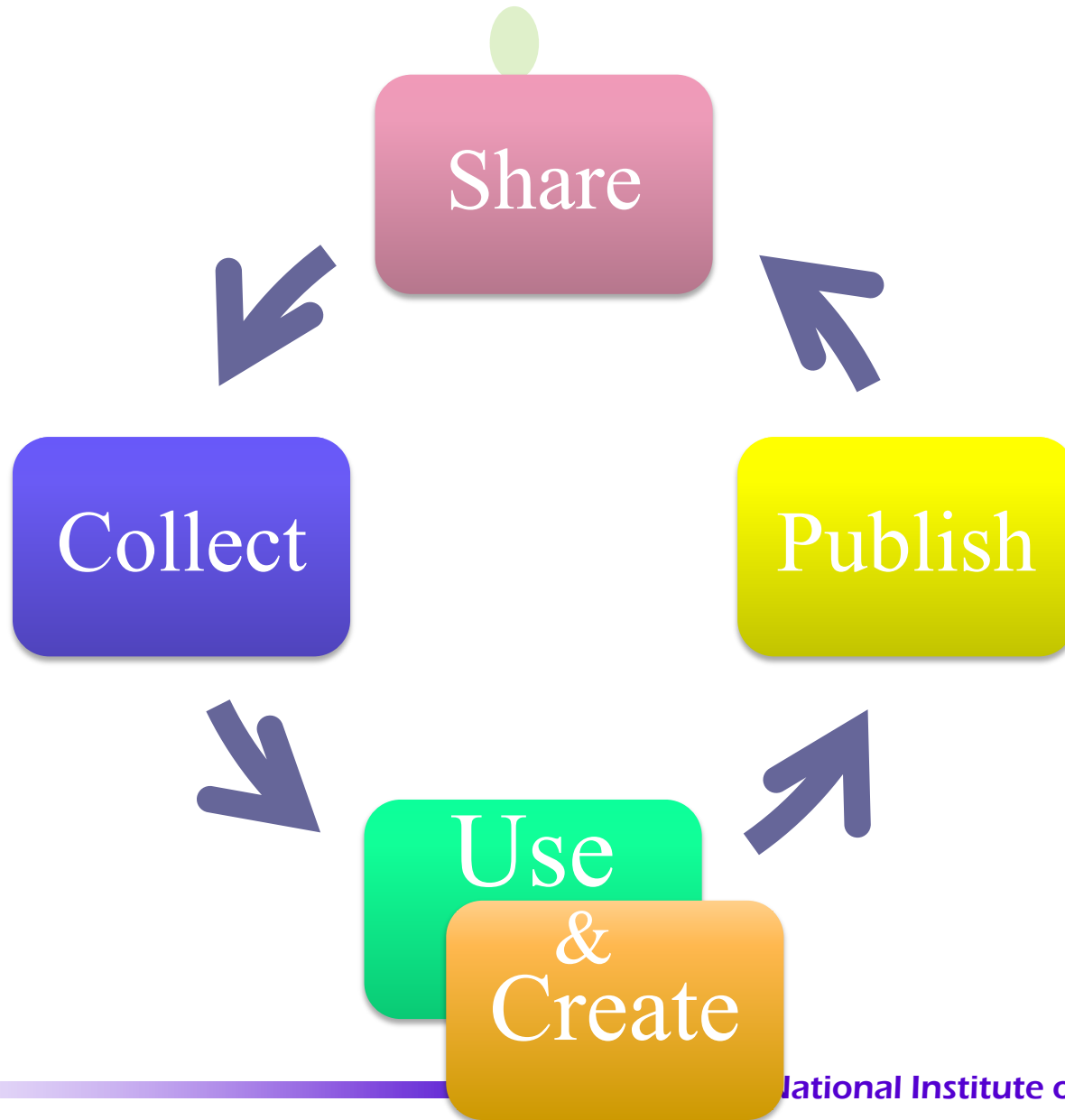
Outline of this course

- Information sharing
 - What is the nature of information sharing. We can learn it from the history of Information in our society. We realize that Information sharing is inevitable.
 - What is the enabling technology for Information Sharing. Information Sharing consists of three layers; the information layer, the semantics layer and the social network layer.
 - In the information layer, we show the basic architecture of information sharing, then pick up some techniques to realize it.
 - In the semantics layer, we introduce **Semantic Web** as a whole, i.e., the basic structure, the theories behind it, the technologies to realize it, and the practice to deploy it.
 - In the social network layer, we show some basic techniques to analyze networks in general.

History of Communication

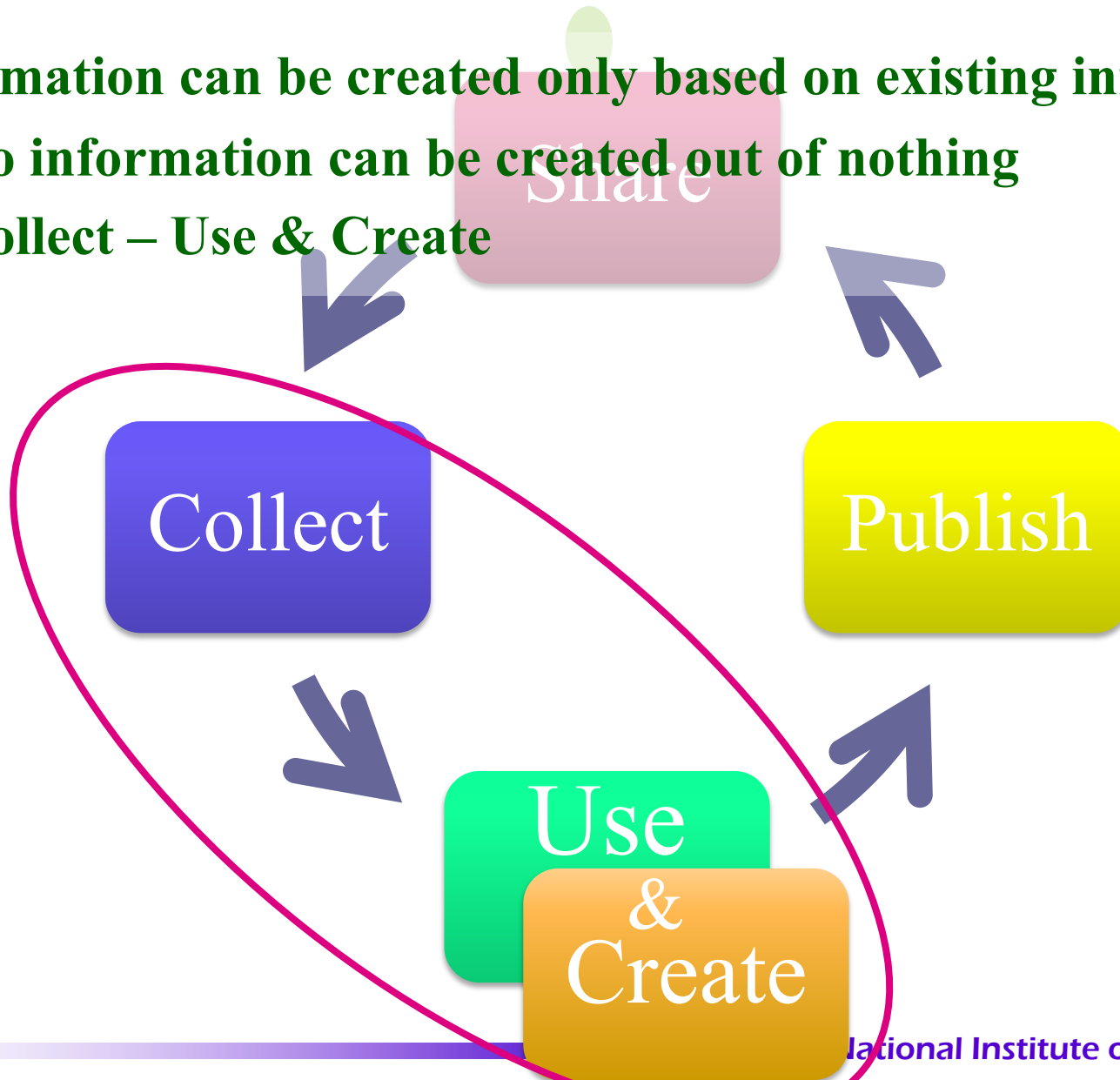
- The path to WWW (web) -

Information Cycle



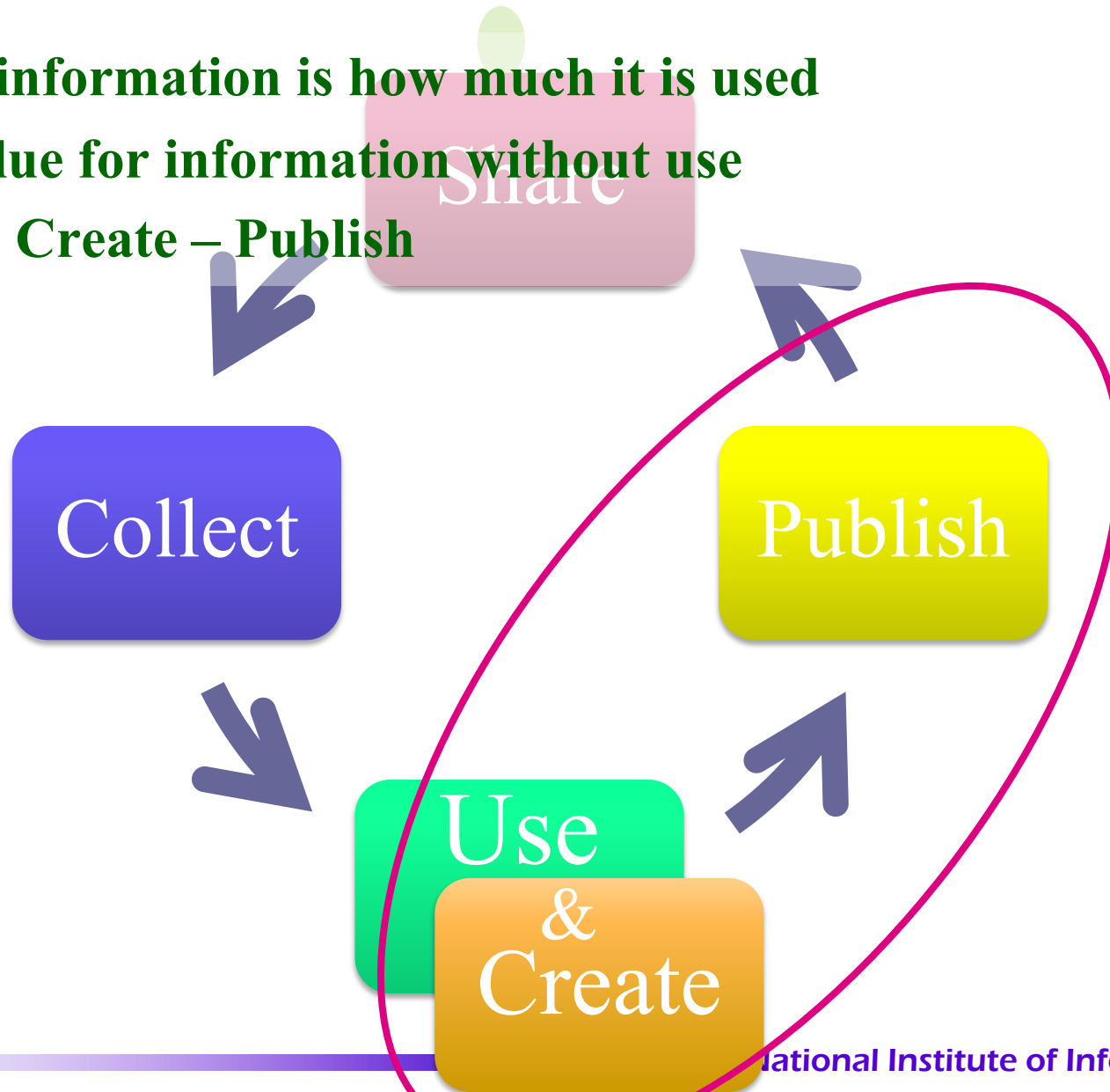
Information Cycle

- Information can be created only based on existing information
 - No information can be created out of nothing
 - Collect – Use & Create

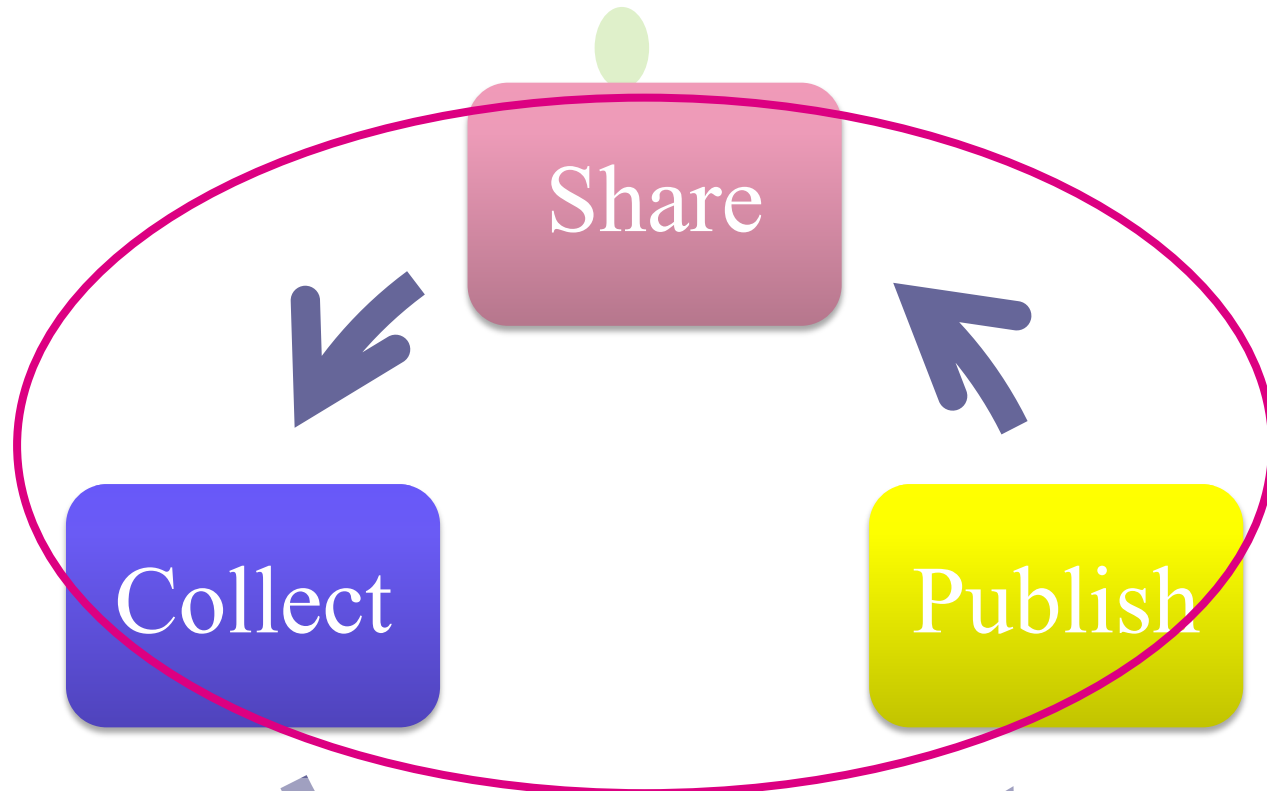


Information Cycle

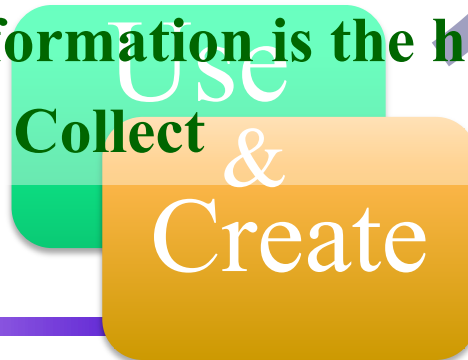
- Value of information is how much it is used
 - No value for information without use
 - Use & Create – Publish



Information Cycle

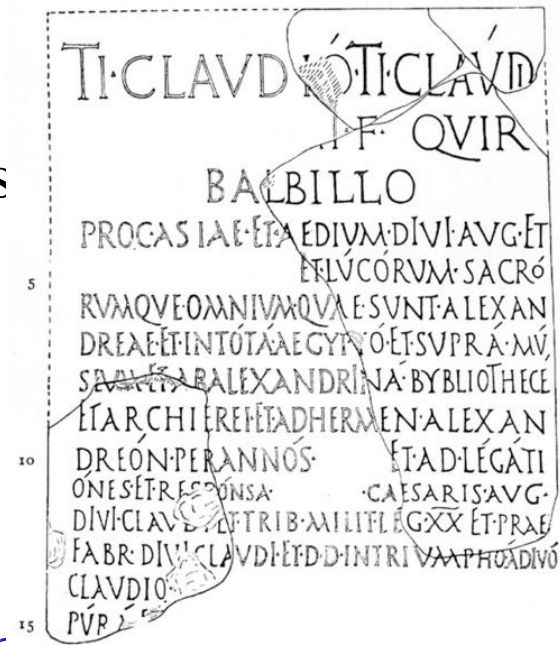


- **Accumulation of information is the wealth of society**
 - **Distribution of information is the health of society**
 - **Publish – Share -- Collect**



Ancient times

- Oral communication
 - Coverage: Everyone
 - Speed of diffusion: very slow
 - Style of diffusion: one-to-one
 - Quantity: small
- Hand writing
 - Coverage: very limited
 - ◆ Aristocracy, monks, bureaucrats, scholars
 - c.f. Library of Alexandria (3c BC)
 - Speed of diffusion: very slow
 - Style of diffusion: one-to-some
 - Quantity: small



Gutenberg Revolution

- Letterpress Printing (活版印刷) 1440
 - Coverage: split-off between authors and readers
 - ◆ Author: professional writers, scholars
 - ◆ Reader: many people
 - Speed of diffusion: faster
 - Style of diffusion: one-to-many
 - Quantity: larger

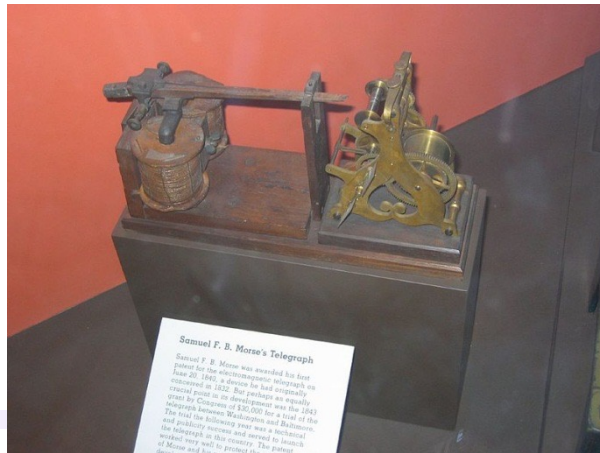


Communication Revolution

- Rapid Growth of communication technology since mid. 19c
 - Telegraph (Samuel Morse ca.1836)
 - Phonograph (Record Player, Thomas Edison ca. 1877)
 - Motion Picture (Cinema, Auguste and Louis Lumière, 1895)
 - Radio Broadcasting (Reginald Fessenden,1906)
 - TV Broadcasting (1928)



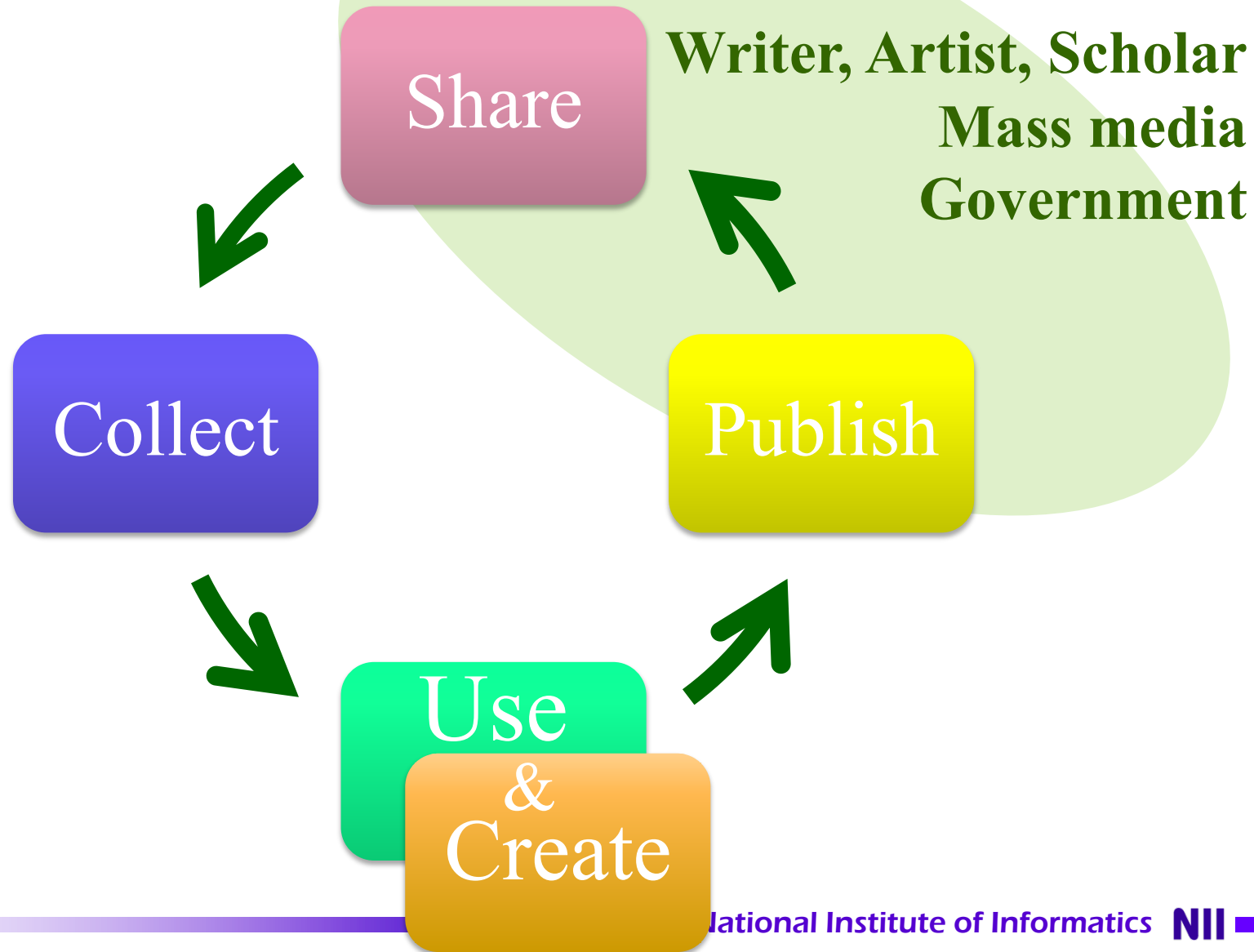
Birth of Mass media



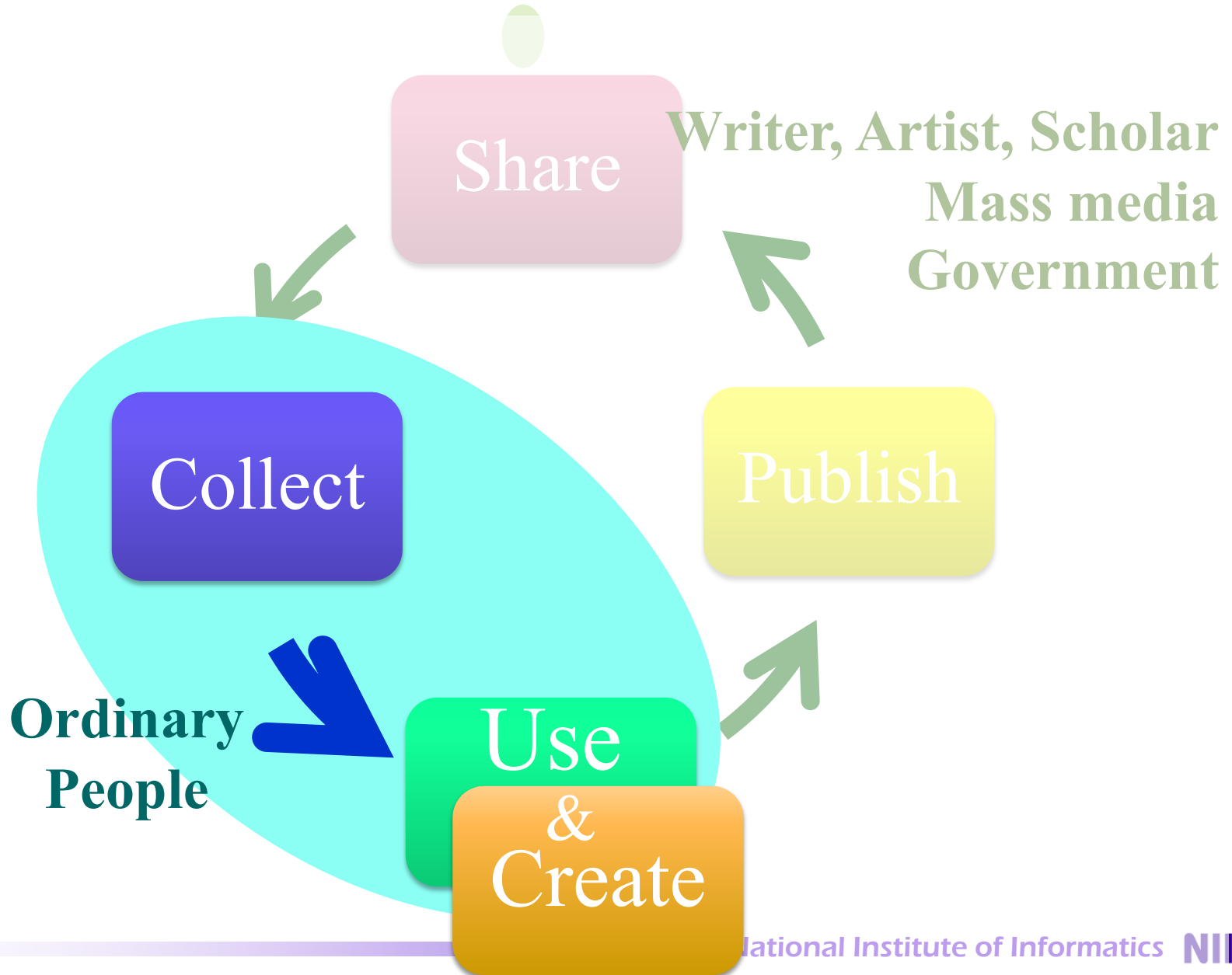
Mass media

- Newspapers, Magazine, Radio broadcasting, TV broadcasting
 - Coverage:
 - ◆ Author: **press**, scholars, politician, government, professional writers
 - ◆ Reader: **Almost everyone**
 - Speed of diffusion: from **simultaneous** to a couple of days
 - Style of diffusion: one-to-so-many
 - Quantity: much larger

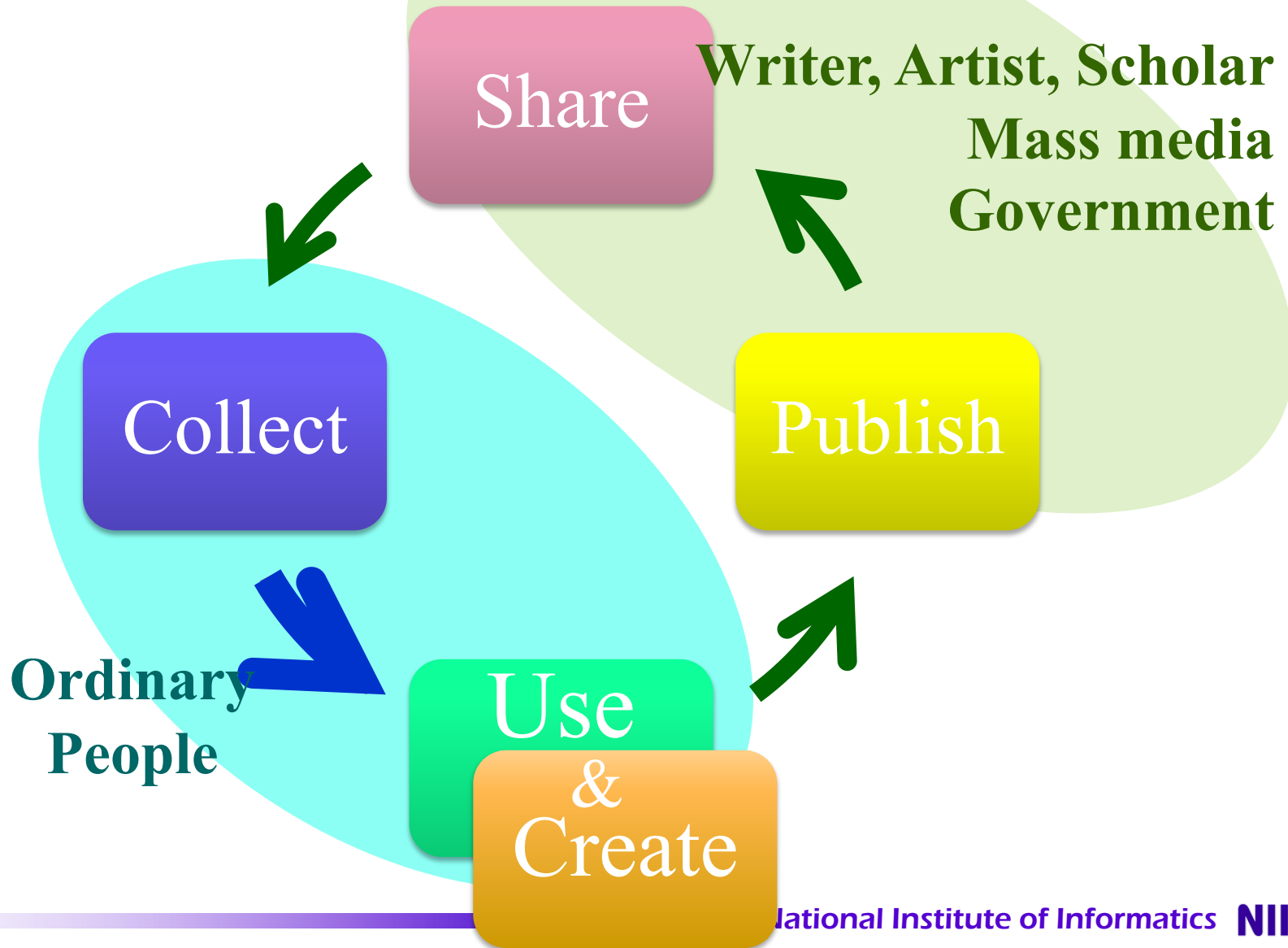
Two social layers on information cycle with Mass Media



Two social layers on information cycle with Mass media



Two social layers on information cycle with Mass Media



Information Cycle with Web

Web

Share

Internet

Open Door to Information Cycle for Ordinary People

Collect

Publish

Search Engine

Web Server

Use

&
Create

Web Browser

HTML Editor

Internet - history

- 1969: First interconnection between UCLA and, SRI
- 1974: Specification of TCP/IP (IETF RFC675)
 - *"Internet" as a single global TCP/IP network*
- 1983: APRANET switched to TCP/IP
- 1984: Domain Name Services started
- 1984: Over 1,000 hosts on Internet
- 1980s: Many interconnection among universities
- 1988: Commercial Internet Providers started in US.
- 1989: Over 10,000 hosts, Countries connected to NSFNET: au, de, il, it, jp, mx, nl, nz, pr, uk
- 1992: Over 100,000 hosts
- 1974: Unix code released
- 1970s: Migration of Unix onto different machines
- 1976: Apple I released
- 1981: IBM PC released
- 1984: Macintosh released
- 1980mid: Unix workstations widely used in universities
- 1991: Windows 3.1

Internet

- The early period: **Infrastructure of personal communication**
 - Protocol: TCP/IP, uucp
 - Usage: Telnet, ftp, email, netnews
 - Functions: very slow, narrow, and unstable TCP/IP connection, no TCP/IP connection (uucp)
- The middle period: **Infrastructure of public communication**
 - Protocol: TCP/IP
 - Usage: email, ftp, netnews, WWW
 - Functions: reliable TCP/IP connection

networking

storage
- Now
 - Protocol: TCP/IP (http)
 - Usage: WWW, SNS, youtube, email
 - Functions: real-time connection

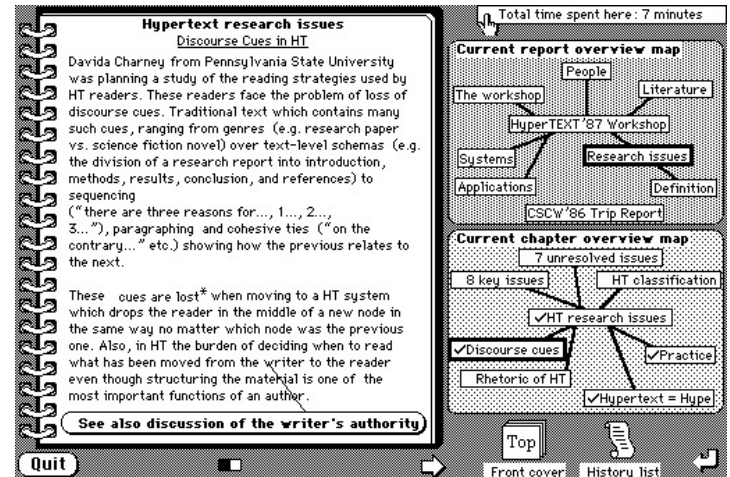
Web (World Wide Web, WWW)

- 1980: Tim Berners-Lee (TBL) developed ENQUIRE
- 1989: TBL wrote a proposal “Information Management: A Proposal”
- 1990: TBL open the first WWW page and HTML browser (editor, called “WorldWideWeb”)
- 1991: TBL posted summary of WWW to alt.hypertext newsgroup
- 1992: National Center for Supercomputing Applications at the Univ. of Illinois at Urbana-Champaign (NCSA-UIUC), led by Marc Andreessen, developed NCSA Mosaic browser (images can be used)
- 1994: TBL founded the World Wide Web Consortium (W3C). *The W3C decided that their standards must be based on royalty-free technology*
- 1996-1998: Commercialization of Web
- 1999-2001: “Dot-com” boom and burst

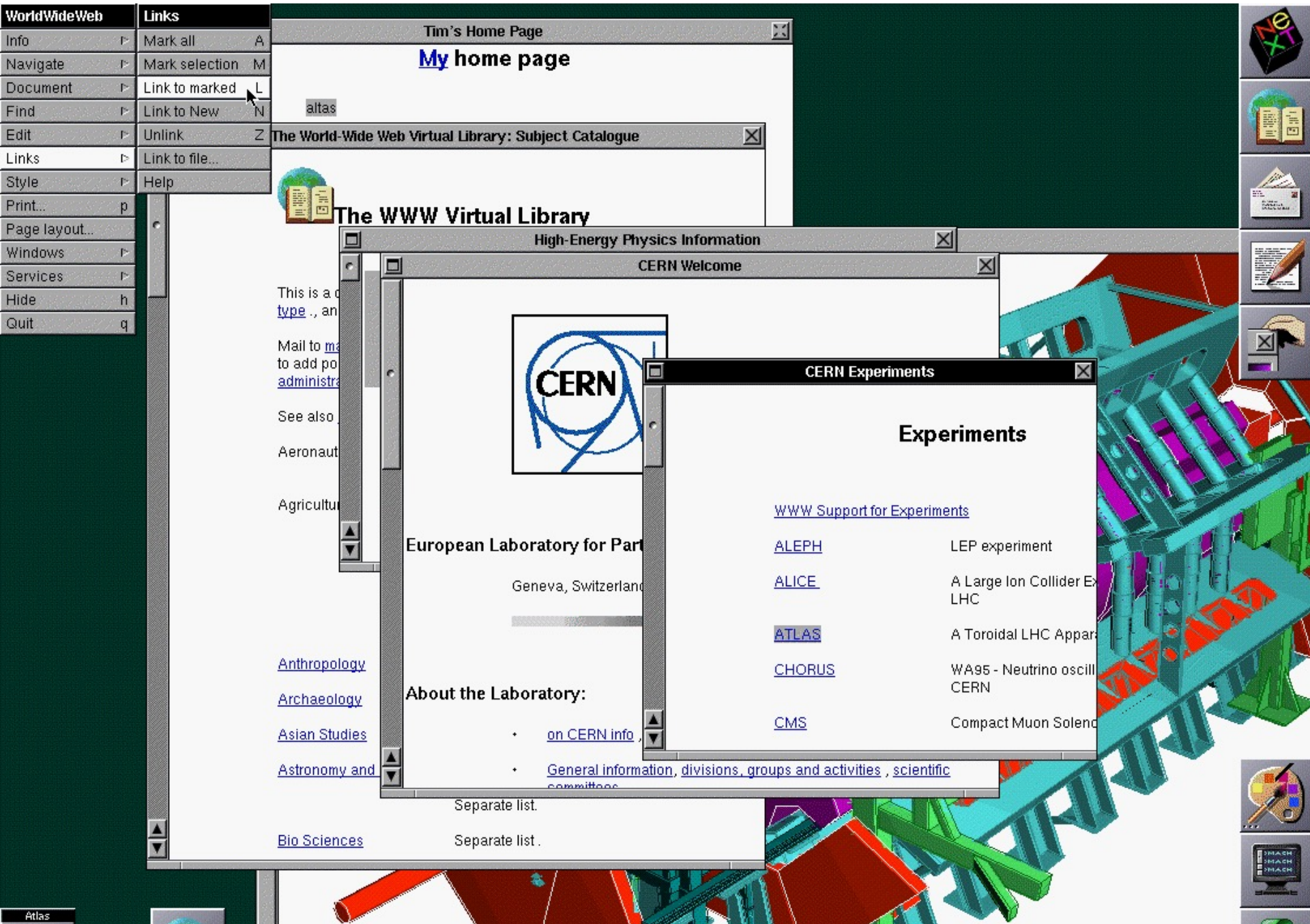
Before Web

- Ancestor Hypertext Systems

- ◆ Memex (V. Bush, 1945)
- ◆ Hypertext Editing System (A. van Dam, T. Nelson, 1969)
- ◆ HyperCard (Apple Computer, 1987)



- ◆ SGML (Standard Generalized Markup Language)



- WorldWideWeb
- Links
- Info ▾ Mark all A
- Navigate ▾ Mark selection M
- Document ▾ Link to marked L
- Find ▾ Link to New N
- Edit ▾ Unlink Z
- Links ▾ Link to file...
- Style ▾ Help
- Print... p
- Page layout...
- Windows ▾
- Services ▾
- Hide h
- Quit q

The World-Wide Web Virtual Library: Subject Catalogue

The WWW Virtual Library

This is a [type](#) ., and

Mail to [mailto:mas](#)
to add po
[administr](#)

See also

Aeronaut

Agricultu

[Anthropology](#)

[Archaeology](#)

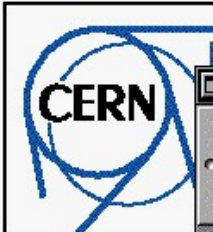
[Asian Studies](#)

[Astronomy and](#)

[Bio Sciences](#)

High-Energy Physics Information

CERN Welcome



European Laboratory for Part

Geneva, Switzerland

About the Laboratory:

- [on CERN info](#)
- [General information, divisions, groups and activities](#) , [scientific committees](#)

CERN Experiments

Experiments

[WWW Support for Experiments](#)

ALEPH	LEP experiment
ALICE	A Large Ion Collider Ex LHC
ATLAS	A Toroidal LHC Appar
CHORUS	WA95 - Neutrino oscill CERN
CMS	Compact Muon Solenoid

World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give
Everything there is online about W3 is linked directly or indirectly to this document, including an [exe](#)

[What's out there?](#)

Pointers to the world's online information, [subjects](#) , [W3 servers](#), etc.

[Help](#)

on the browser you are using

[Software Products](#)

A list of W3 project components and their current state. (e.g. [Line Mode](#) ,[X11 Viola](#) , [NeXT](#)

[Technical](#)

Details of protocols, formats, program internals etc

[Bibliography](#)

Paper documentation on W3 and references.

[People](#)

A list of some people involved in the project.

[History](#)

A summary of the history of the project.

[How can I help ?](#)

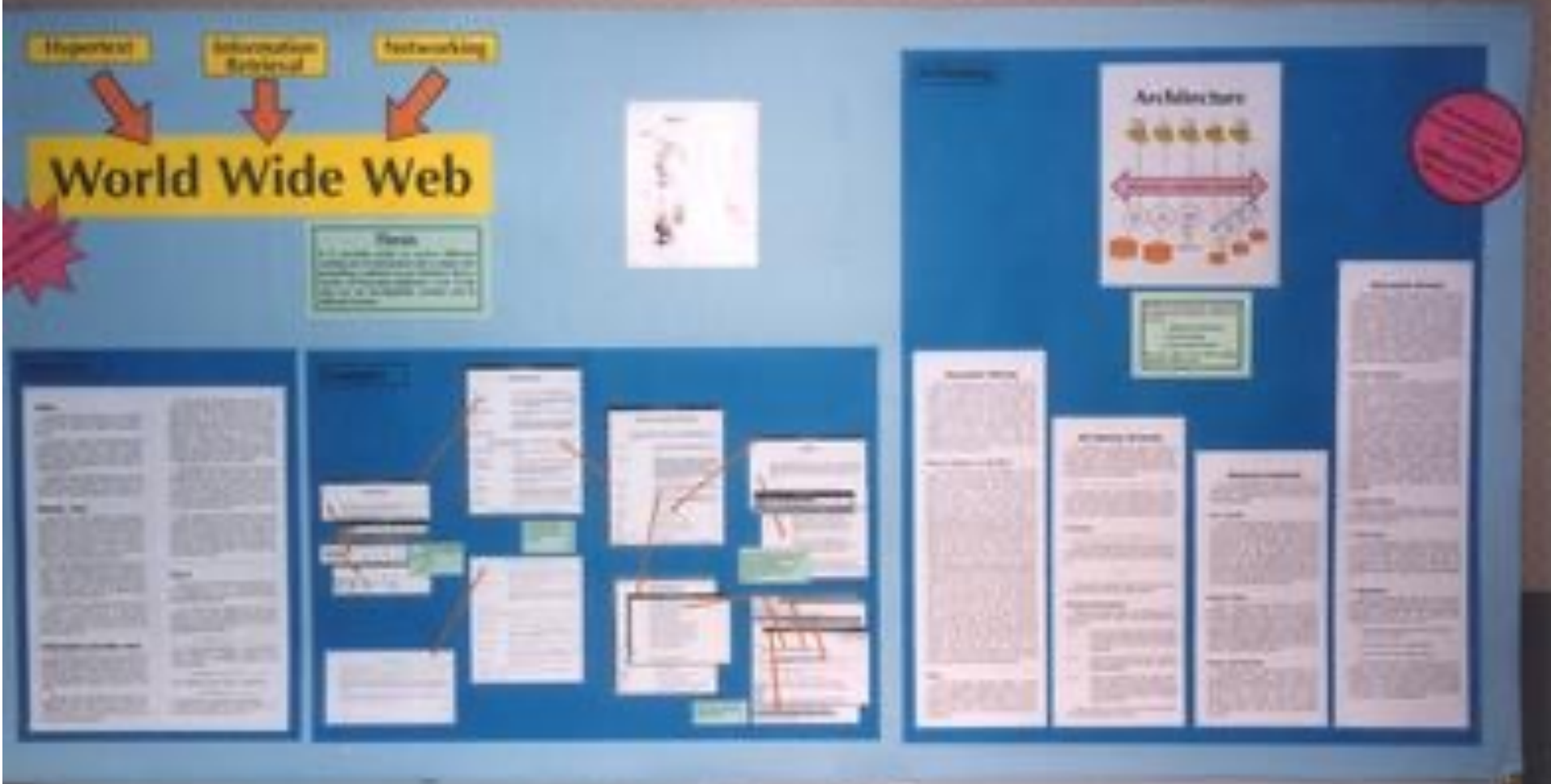




話すTim Berners-Lee (Hyderabad, India, 2011)

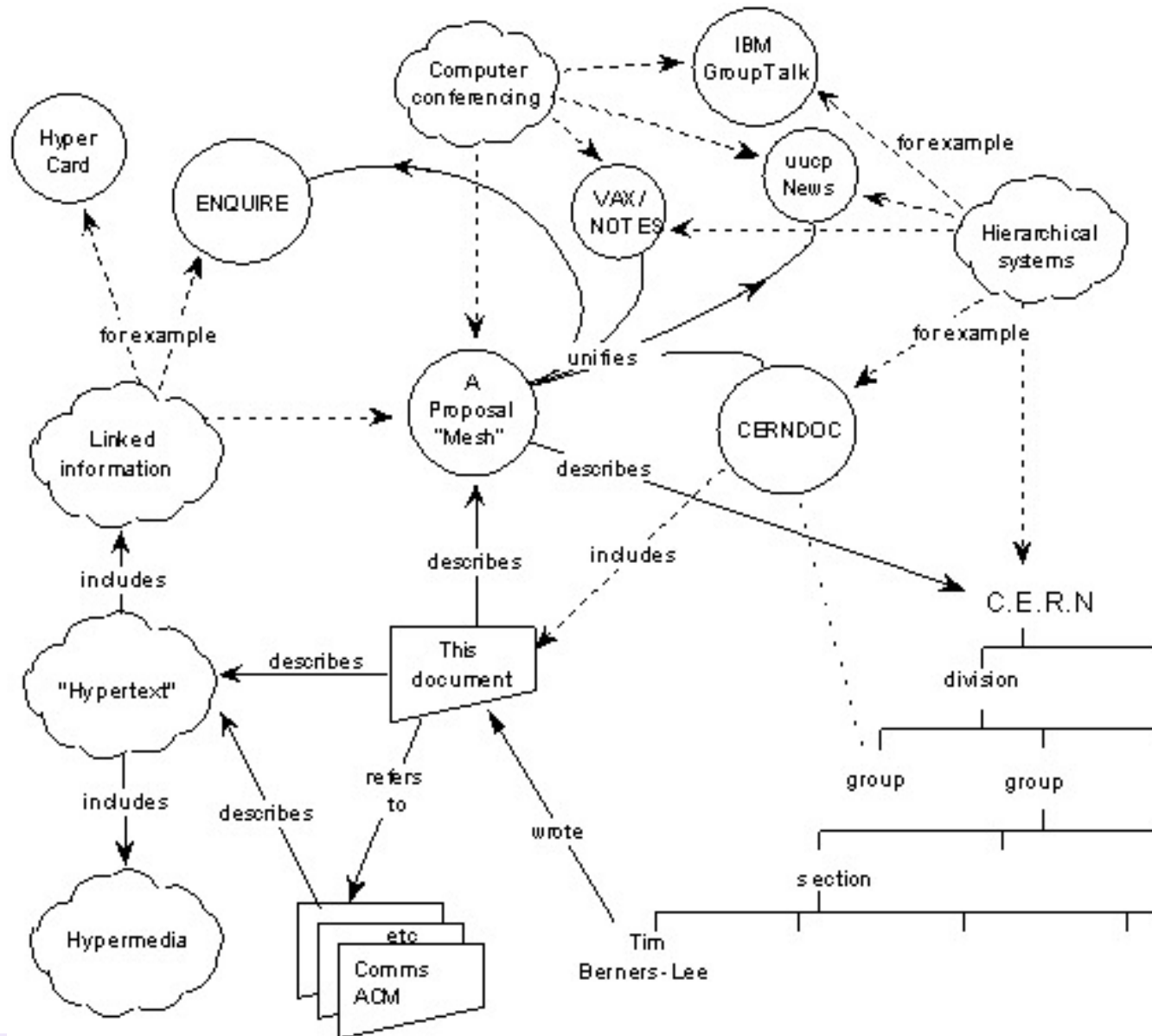


踊るTim Berners-Lee (Banff, Canada, 2007)



Hypertext '91

Information Management: A Proposal



File Options Navigate Annotate Help

Document Title: Sample Form

Document URL: http://192.168.0.1/examples/sample-form.html

Sample Form

Text field: xxx

Text area: Initial data

Password: ***

Hidden field:


Radio buttons: A B C

Check boxes: 1 2 3 4 5

Selection: option 1

Reset buttons: Reset CLEAR

Submission buttons: Submit Query SEND

Image button: 

Back Forward Home Reload Open... Save As... Clone New Window

NCSA Mosaic

Yahoo! - NCSA Mosaic

File Edit History Manager View Navigate Tools Hotlists Help

http://www.yahoo.com

What's New Check Email Personalize Help

Yahoo! Auctions
Pokemon, Rolex, 'N Sync

free_email@yahoo.com

Win Free Flowers
FTD.COM

Search advanced search

Shopping - Auctions - Yellow Pages - People Search - Maps - Travel - Classifieds - Personals
Games - Chat - Clubs

Mail - Calendar - Messenger - Companion - My Yahoo! - News - Sports - Weather - TV

Tue 06/06/00 9:33:04

Web History - Web Architecture and Information Management
 Spring 2010 — INFO 190-02 (CCN 42509)
 Wilde and Ryan Shaw, UC Berkeley School of Information, 2010-01-25
<http://dret.net/lectures/web-spring10/history>



武田 英明

- [略歴](#)
- [大学の公式ページにおける個人ページ](#)
- [主な研究課題](#)
- [研究発表](#)
- [研究関連情報](#)
 - [インターネットからの情報獲得と統合](#)
 - [知識を持つ環境研究ページ](#)
 - [設計研究ページ](#)
 - [国内の設計関連研究者ページ](#)
- [連絡先](#)
 - **住所** 〒630-01 生駒市高山町8916-5 奈良先端科学技術大学院大学情報科学研究科
 - **部屋番号** B0x(情報科学研究科 B・E階)
 - **電話**: 0743-72-XXXX
 - **電話**: 0743-72-XXXX(学科事務室)
 - **ファックス**: 0743-72-XXXX
 - **E-mail**: takeda@is.aist-nara.ac.jp
- [その他](#)

[西田研究室](#) / [情報科学研究科](#) / [奈良先端科学技術大学院大学](#)

図2 1996年ごろの研究者のWebページ例

Web Architecture

- Essential technologies
 - Uniform Resource Identifier (URI): a system of globally unique identifiers for resources on the Web and elsewhere
 - HyperText Markup Language (HTML): the publishing language;
 - Hypertext Transfer Protocol (HTTP)
- Server-Client Architecture base on URI, HTML, HTTP
 - Fully distributed systems
 - ◆ No center needed
 - ◆ How to access?
- Royalty-free policy
 - Rapid software development (everyone can develop software)
 - Rapid software deployment (everyone can use software)

Some statistics (but old!)

- Over 20 billion pages (google, Yahoo!)

- 2006 Feb: Yahoo!

- Over 80 million servers (80,655,992)

- 2006 April

- Web Server Survey

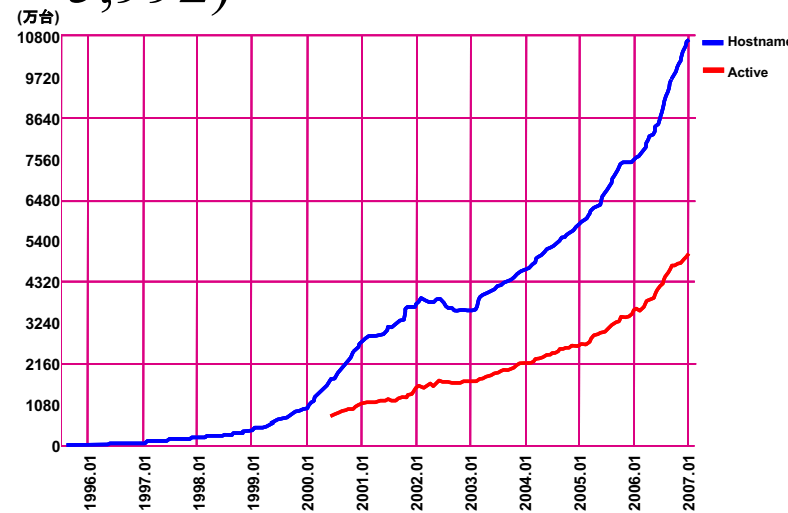
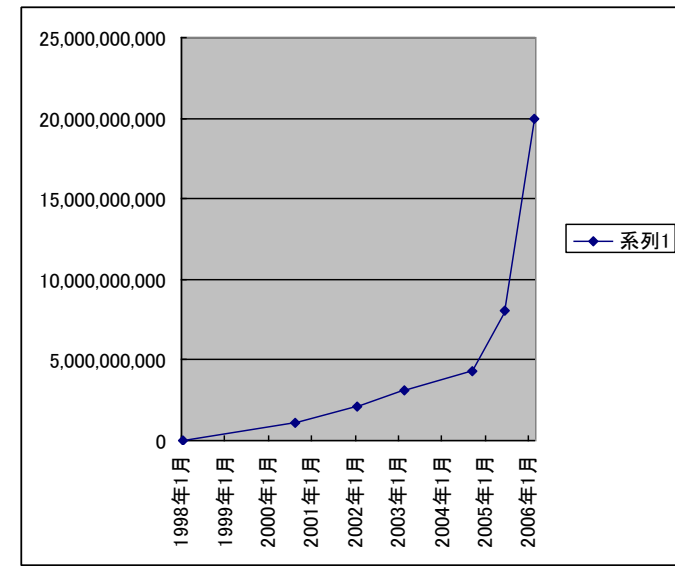


図1: Webサーバの台数の変化 (NetCraft, <http://www.netcraft.com>)

- 0.69 billion users (14% of population)

- 2006 March

- comScore <http://www.comscore.com/press/release.asp?press=849>

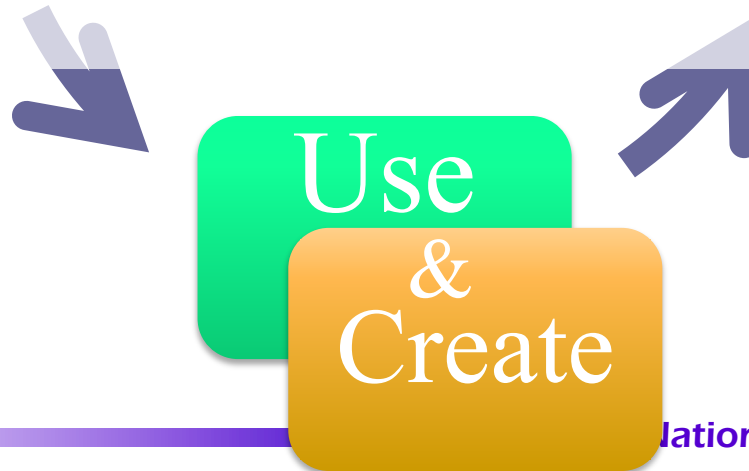
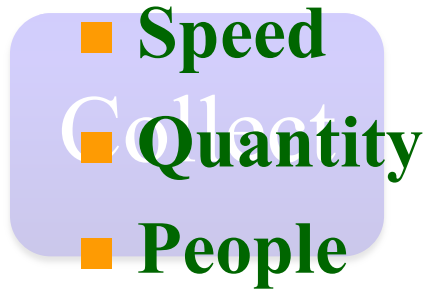
Top 15 Online Populations by Country, Among Visitors Age 15+* March 2006	
Unique Visitors (000)	
Source: comScore World Metrix	
	Unique Visitors (000)
Worldwide Total	694,260
United States	152,046
China	74,727
Japan	52,100
Germany	31,813
United Kingdom	30,190
South Korea	24,645
France	23,884
Canada	18,996
Italy	16,834
India	16,713
Brazil	13,186
Spain	12,452
Netherlands	10,969
Russia	10,833
Australia	9,735

Information Cycle

Web



- **Web accelerate Information Cycle in**



Why is Web diffused so much?

- Openness
- Economics
- Easiness

Why is Web diffused so much? - Openness

- Openness:
 - Open to participate
 - Open to relate
- Internet
 - Open to participate
 - ◆ Open to join the computers to the network
 - Open to relate
 - ◆ Open to communicate to any computers
- Web
 - Open to participate
 - ◆ Open to make web pages
 - Open to relate
 - ◆ Open to make links to others

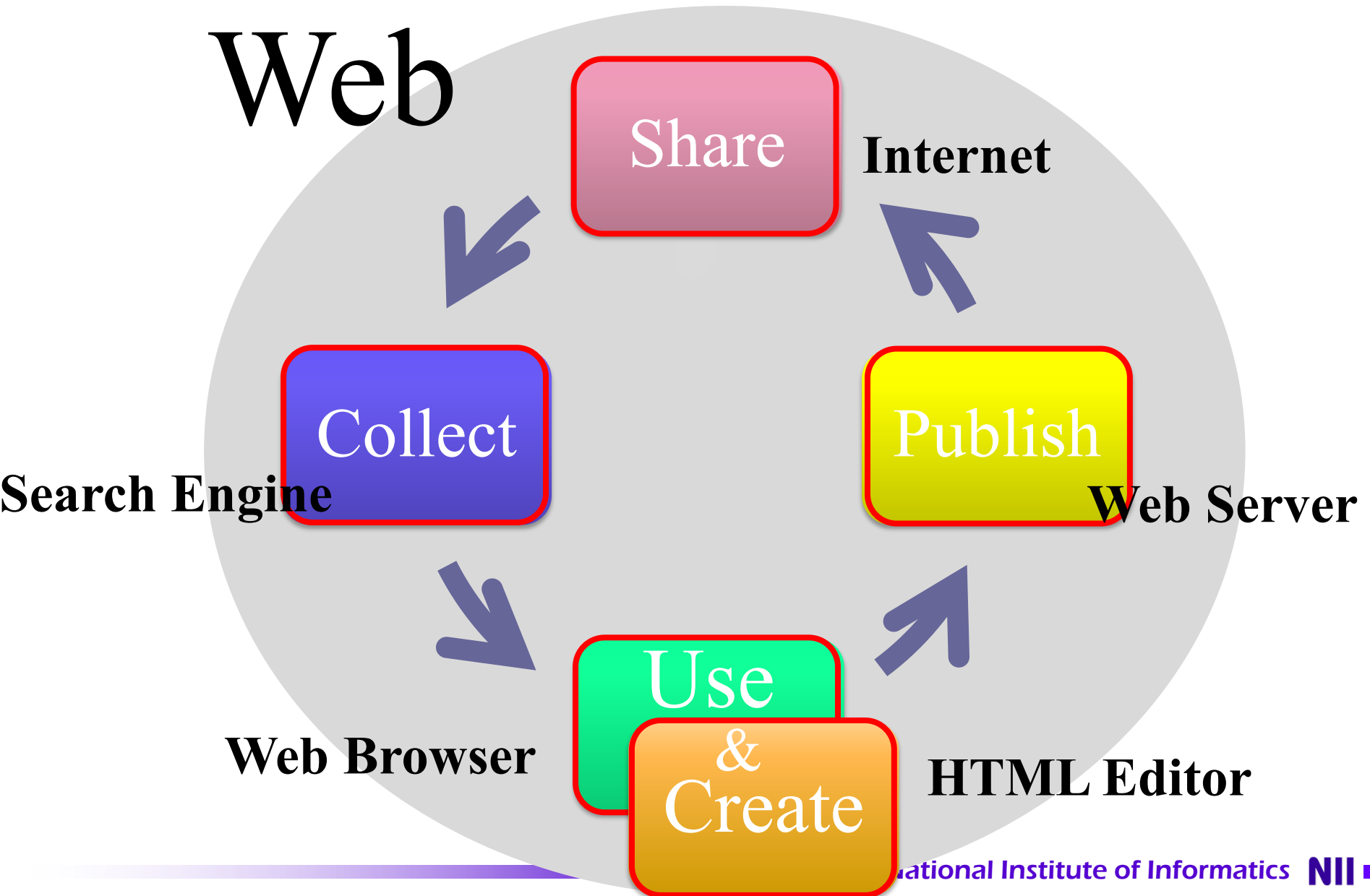
Why is Web diffused so much? - Economics

- Cost is not proportional to quantity of information
 - Packet communication (TCP/IP)
- Cost is proportional to quantity of information in legacy media
 - For people: A really new media that people can use

Why is Web diffused so much? - Easiness

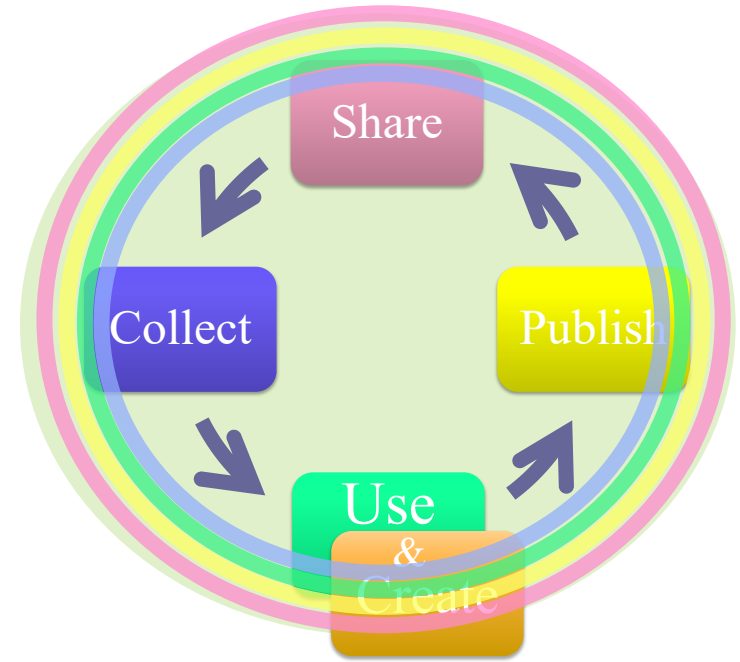
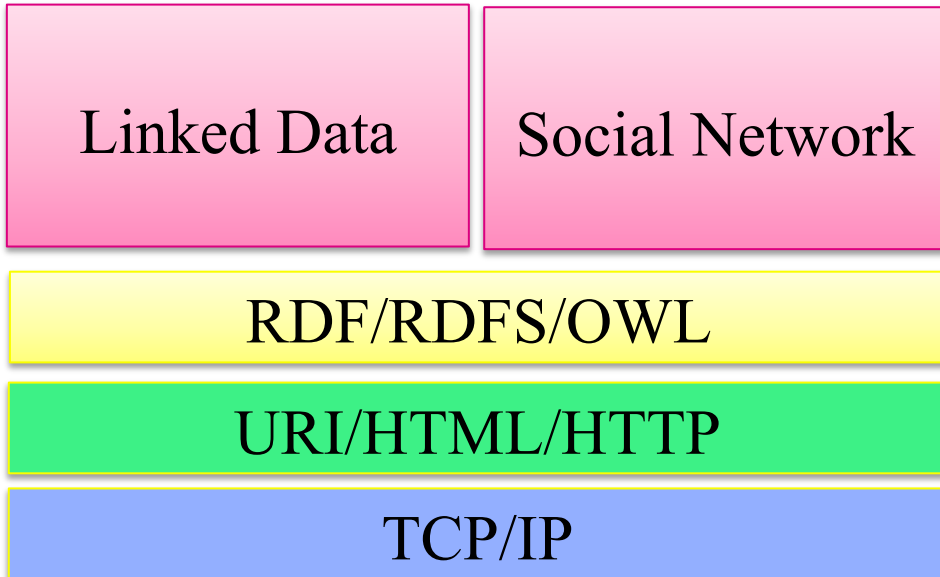
- The first computer systems for ordinary people
- Internet is a system for computer professionals (or *geeks*)
 - Read manuals, solve problems by yourself
- TBL designed Web not for computer professionals but for researchers (physics)
 - Simplicity
 - Robustness

Information Cycle with Web



Layers for Information Cycle

- Internet
- Web
- **Semantic Web**
- Linked Data
- Social Network



Semantic Web

Web

Internet