Design Considerations for a Distributed Information System for Mathematics in Germany

Joachim Lügger

luegger@zib-berlin.de

Konrad-Zuse-Zentrum
für Informationstechnik Berlin (ZIB)

Distributed Information System (DIS)
- Supplements and Comments -

Information Coordinator
- coordinates within Department, Institution
  (students, secretaries, librarians)
- cooperates with uni-library, computing center

Departmental Systems, incl. Libraries
- local literature (catalog of books, journals etc.)
- rare special literature (e.g. special workshops)
- uniform database access (window oriented)?

Living Mathematical Museum
- ancient part (historians)
- modern part (multi-media)
- experimental laboratory (didactically)

Document Delivery and Electronic Publishing via CD-ROM
- via TIB Hannover, SUB Göttingen
- via university libraries (e.g. within Nordrhein Westfalen)

Integration of "Zentralblatt" - Cooperation with FIZ Karlsruhe
- electronic reviewing (speeding up, cost reduction)
- integration of services (electronic offers)
- close cooperation with publishers

Central Support for Small Institutions
- central information store (high speed access)
- information distribution service for publishing houses

Cooperation, Participation in Standardization Activities
- national: GI, DPG, DGCH, ...,Internet Society,...
- international: AMS, EMC, ..., Euromath, ...
- with other faculties (within universities)

Further Suggestions and Comments are Appreciated

DIS Tasks and Pilotprojects

I. Scientific/Technical

- Set up of Electronic Information Systems in Mathematics
- Installation of Information Servers and Clients at all Partners
- Distributed Offer of Preprints and Lecture Notes
- Distributed Offer of Software and Data Collections
- Incorporation of Mathematical Information Systems Worldwide
- Access to Electronic Library Catalogs and Databases
- Organization and Creation of Electronic Mathematical Journals
- Electronic Offer of Scanned Historical Documents and Books
- Organization of a Living Mathematical Museum
- Gaining Experiences in Electronic Publishing via CD-ROM
- Incorporation of Electronic Document Delivery
- Providing Central Support for Small Institutions and Enterprises
- Creation of a Framework for Various Kinds of Electronic Reviewing
- Electronic Project Organization

II. Organizational Infrastructure

- Information Coordinators
- Forum for Mathematical Information
- Integration of Special Tasks Regarding "Zentralblatt"
- Cooperation with Libraries, Publishers, Industry, other Faculties
- National and International Cooperation, Standardization Activities
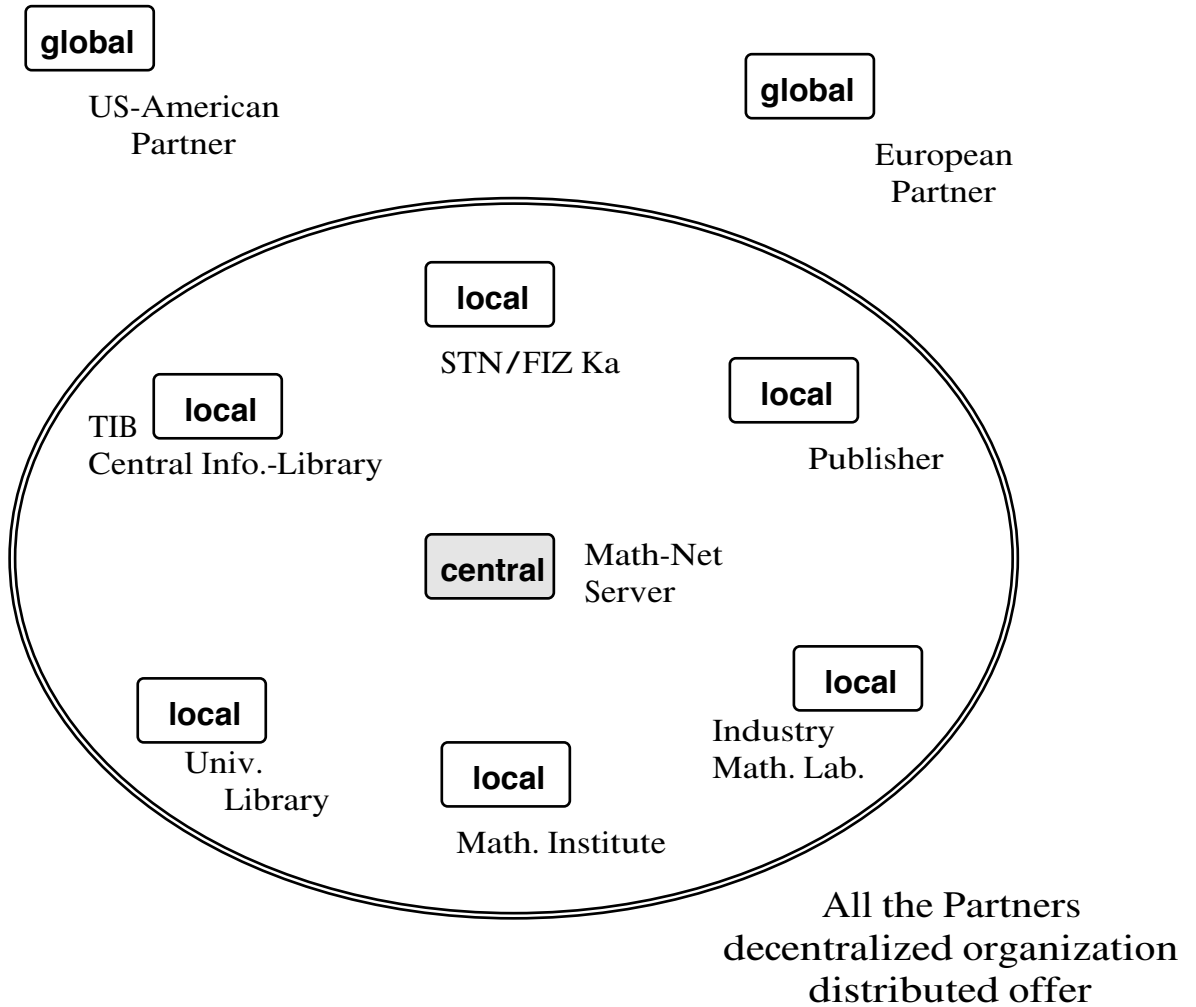
Under the Auspices of DMV
DMV WG on Information and Communication

Distributed Information Collections

Information	Primary	Reference	Class
Publications	Preprints, Reports Books Lecture Notes, Scripts E-Journals	Title Abstract/Review Bibl. Reference Link	MSC
Software	Programs Program Libraries Software Systems Testdata Data Collections	Title Index/Readme Data Sheet Review Link	GAMS
Info-Systems	Home Pages Top Menus ftp-Path telnet address	Title Short Description Readme Link	?
Organization	Departments Institutes, Libraries Projects Seminars Courses Services, Facilities Chairs, Members	Short Description Access Ref. Prospectus Title Tour Map Link	MSC?
Math Museum	Ancient Docs Ancient Books History Historic Devices Mathematicians Modern Math Arts Modern Experiments Didactic Laboratory Interactive Books Computers	Title Abstract Time Table Photo-Ref. Short Biography Icon/Graphics Form Video-Ref. Graphic-Ref. Link	?
Persons	Staff Post Docs Guests Research Students	WhitePage Nickname E-mail Address Interest	?

Decentralized vs. Centralized

Global vs. Local
Decentralized vs. Centralized

global

US-American
Partner

global

European
Partner

local

STN/FIZ Ka

local

TIB
Central Info.-Library

local

Publisher

central   Math-Net
Server

local

Industry
Math. Lab.

local

Univ.
Library

local

Math. Institute

All the Partners
decentralized organization
distributed offer

Characteristics

⇒ Large Set of Items

- local: 100 - 1000
- national: > 5000
- global: unlimited

⇒ Complexity

- web structure ("outside" of docs)
- doc structure ("inside")
- classification ("ordering" of docs)

⇒ Highly Dynamic

- Large Groups of Contributors
- weakly coordinated
- varying levels of experience

Problems
- Best Structure? - "outside"
- Unifying Structure? - "inside"
- Best View? - "ordering"

Depending on
- *the problem type you have*
- *your special orientation, experiences and skills*

This may change from time to time (author & reader)

Distributed Storage Schemes
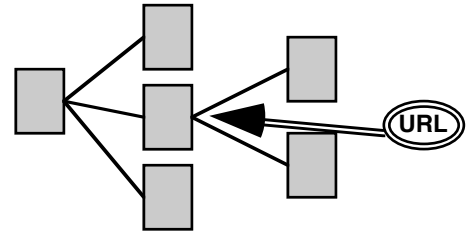
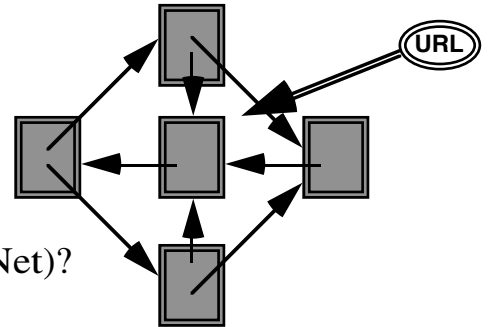|		
Publications	ftp	
Software	(Archie)	

Universal Resource Locator — URL

Info-Systems	gopher
	(Veronica)

Organization	HyperText
Museum	• WWW
	• Hyper-G
	• Mosaic
	+ Search Index

Persons	WhitePage+whois (Opt-Net)?
	netfind?
	x.500?

Libraries	Special Database
	+ Retrieval
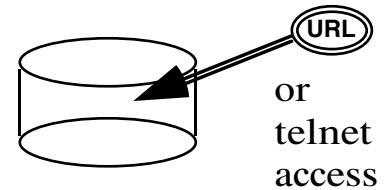
All Kinds of Protocols, Formats and Types

or telnet access

Text, ASCII	Postscript, DVI
TeX, LaTeX	TIFF
MATH, MR	JPEG
BibTeX	MPEG
SGML, HTML	CD-ROM Formats
Library Structures ...	
WhitePages	

Open with respect to the Internet

*All types should also be reflected via the central parts
of the distributed information system for mathematics.*

Hierarchy of Collections

Math

Org Softw InfoS Publ Mus Pers Conf

. . .   .

Dep Libr Serv ProjL   Prepr Books E-Jour Rev   DIS Zbl Opt

. . .   .

Inst Sem Cours   Elec Paper   open closed

. . .   .

Chair Memb Proj Fac

. . .   .
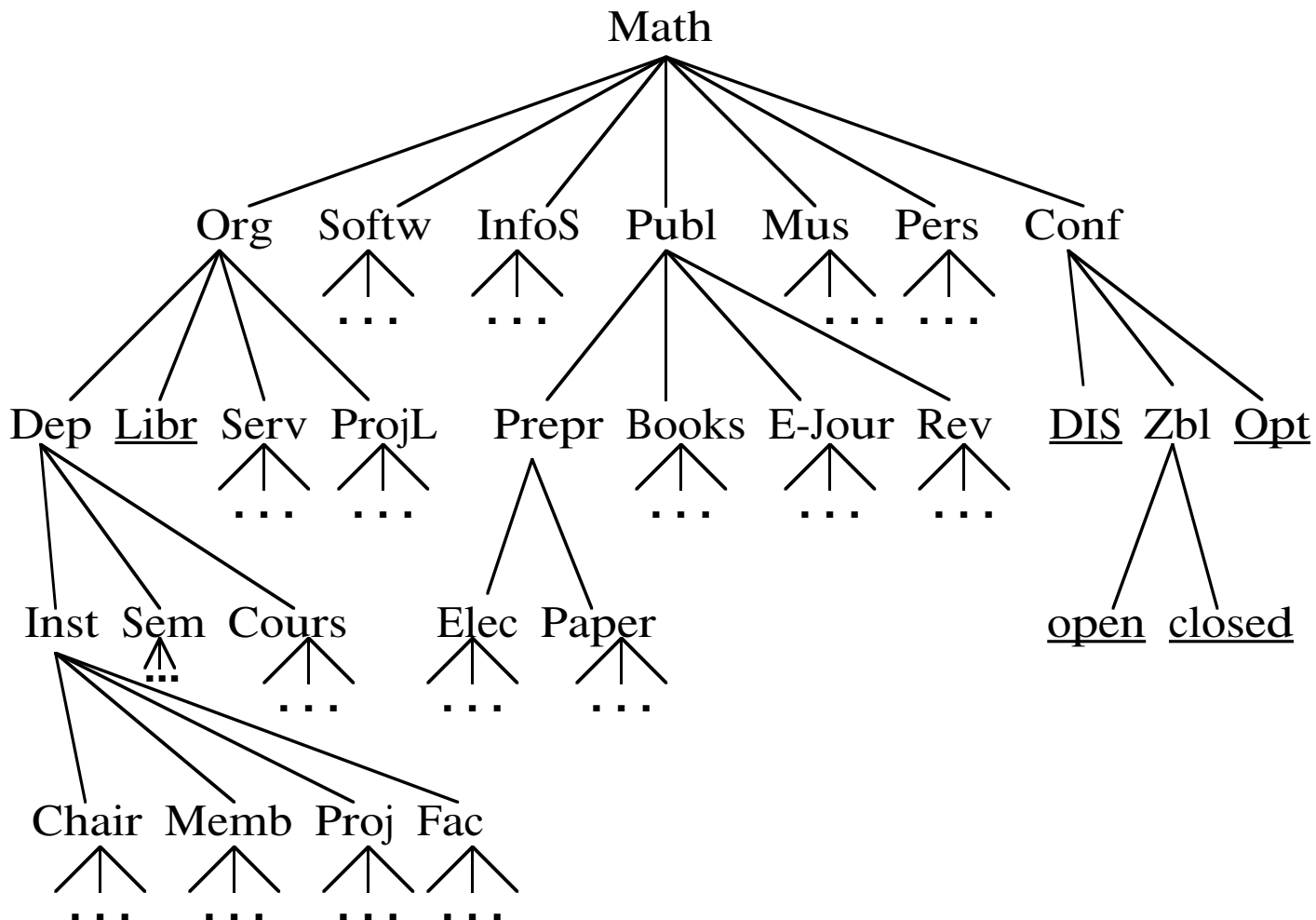
Single and Uniform Structure?
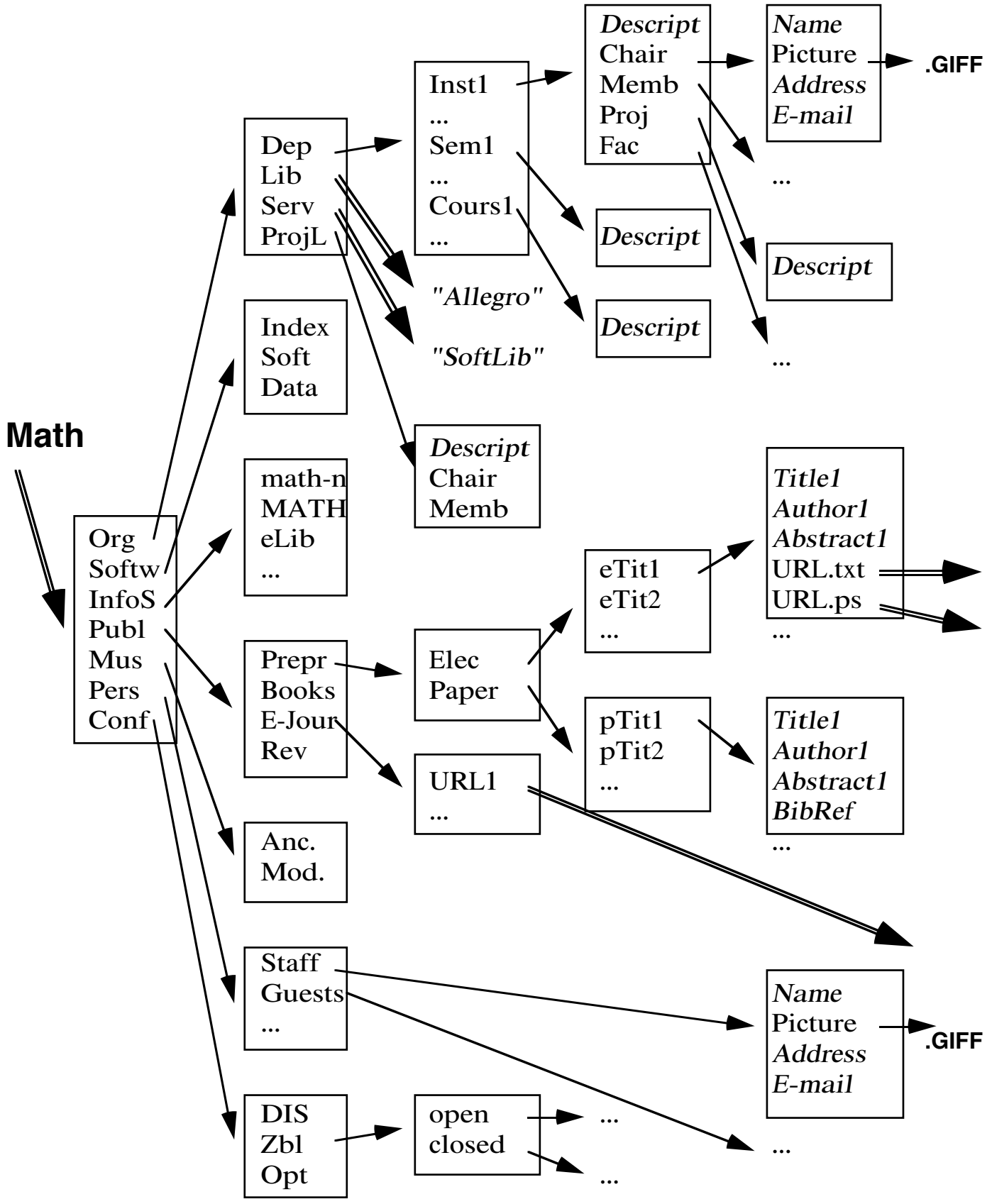
- *decentral   vs. central*
- *large   vs. small*
- *academic   vs. publisher   vs. info-center*

Compatibility within Organization (e.g. University)

Easy Use by Hypertext

A Possible Hierarchic Hypertext Structure

Math

Org	
Softw	
InfoS	
Publ	
Mus	
Pers	
Conf	

Dep
Lib
Serv
ProjL

Index
Soft
Data

math-n
MATHe
eLib
...

Prepr
Books
E-Jour
Rev

Anc.
Mod.

Staff
Guests
...

DIS
Zbl
Opt

Inst1
...
Sem1
...
Cours1
...

"Allegro"

"SoftLib"

Descript
Chair
Memb
Proj
Fac

Descript

Descript

Descript
Chair
Memb

Name
Picture
Address
E-mail

.GIFF

...

Descript

...

Elec
Paper

URL1
...

eTit1
eTit2
...

pTit1
pTit2
...

Title1
Author1
Abstract1
URL.txt
URL.ps
...

Title1
Author1
Abstract1
BibRef
...

open
closed

...

...

Name
Picture
Address
E-mail

.GIFF

...

Search Strategies in Hypertext Systems

Problems for Readers

- How to find something?
- How to find it again?
- How to find it efficiently?
- Every information item seen?

"Lost in Hyperspace Syndrom"   (mild form)

The WWW + WAIS Solution

- WAIS-index all documents   -   Full text indexing
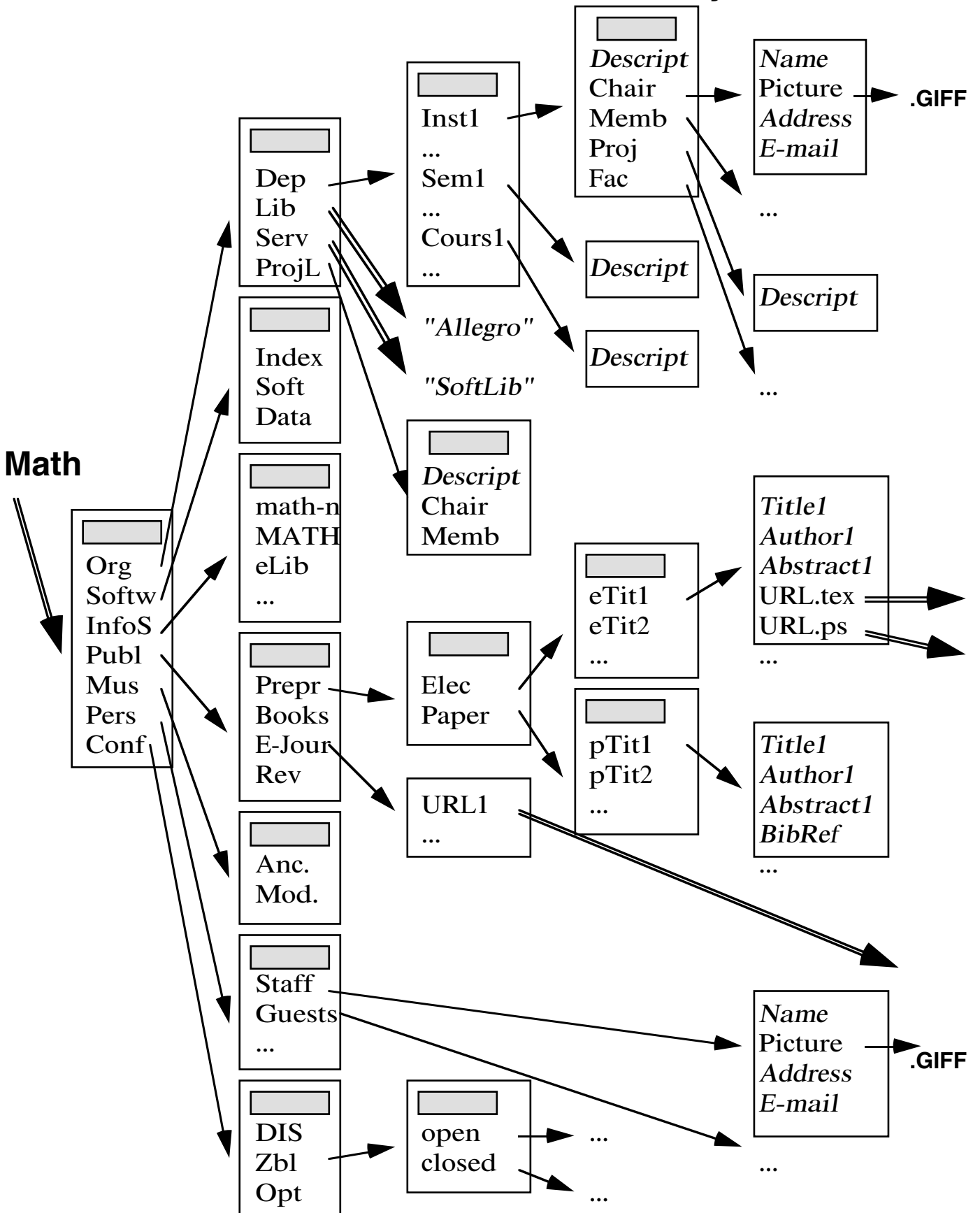- Combine WAIS- and WWW-Server

"Flat WAIS Index Query" possible via Mosaic

The Hyper-G Solution

- Titles and texts of documents are indexed automatically
- they can be queried via the Harmony-Browser
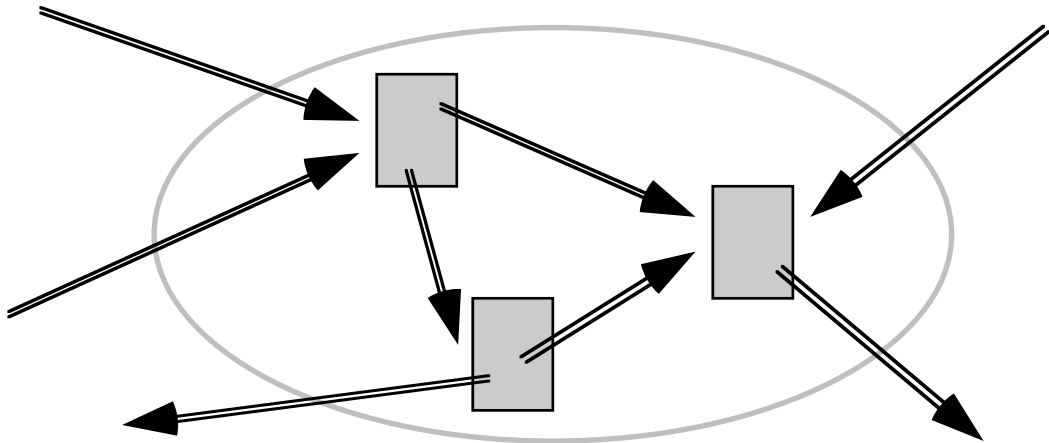- there is a WWW to Hyper-G gateway

"Hierarchic Query" possible via Mosaic

<u>Collection Oriented Hierarchic Query</u>

Math

Org
Softw
InfoS
Publ
Mus
Pers
Conf

Dep
Lib
Serv
ProjL

Index
Soft
Data

math-n
MATHE
eLib
...

Prepr
Books
E-Jour
Rev

Anc.
Mod.

Staff
Guests
...

DIS
Zbl
Opt

Inst1
...
Sem1
...
Cours1
...

"Allegro"

"SoftLib"

Descript
Chair
Memb
Proj
Fac

Descript

Descript

Descript
Chair
Memb

Name
Picture
Address
E-mail

.GIFF

...

Elec
Paper

URL1
...

eTit1
eTit2
...

pTit1
pTit2
...

Title1
Author1
Abstract1
URL.tex
URL.ps
...

Title1
Author1
Abstract1
BibRef
...

open
closed

...

...

Name
Picture
Address
E-mail

.GIFF

...

Openness   vs.   Integrity

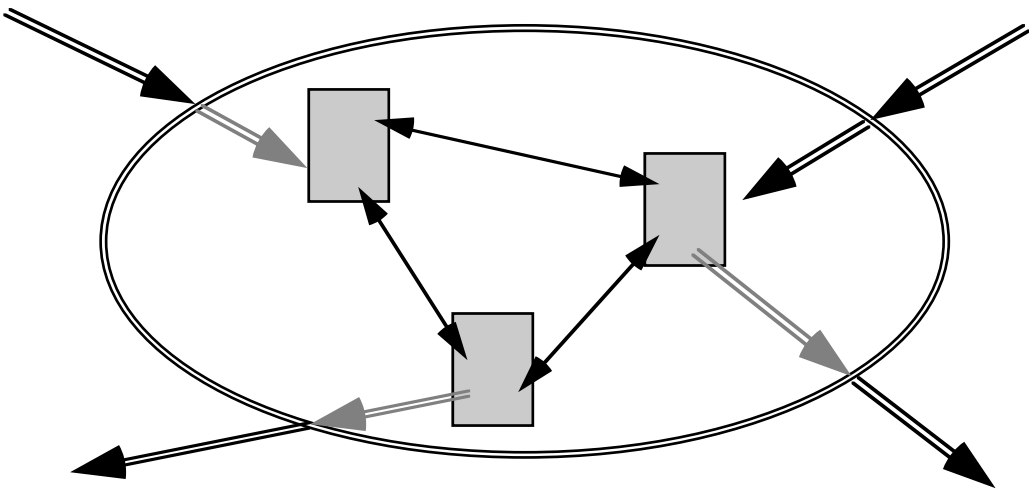I. WWW uses URLs without exceptions - burned into HTML docs

Problem: moving/deleting "not possible" any more
- Hierarchic file structure is fixed
- How long will an URL be valid?
  What's wrong with URLs?

II. Hyper-G uses bidirectional links - maintained by a link server

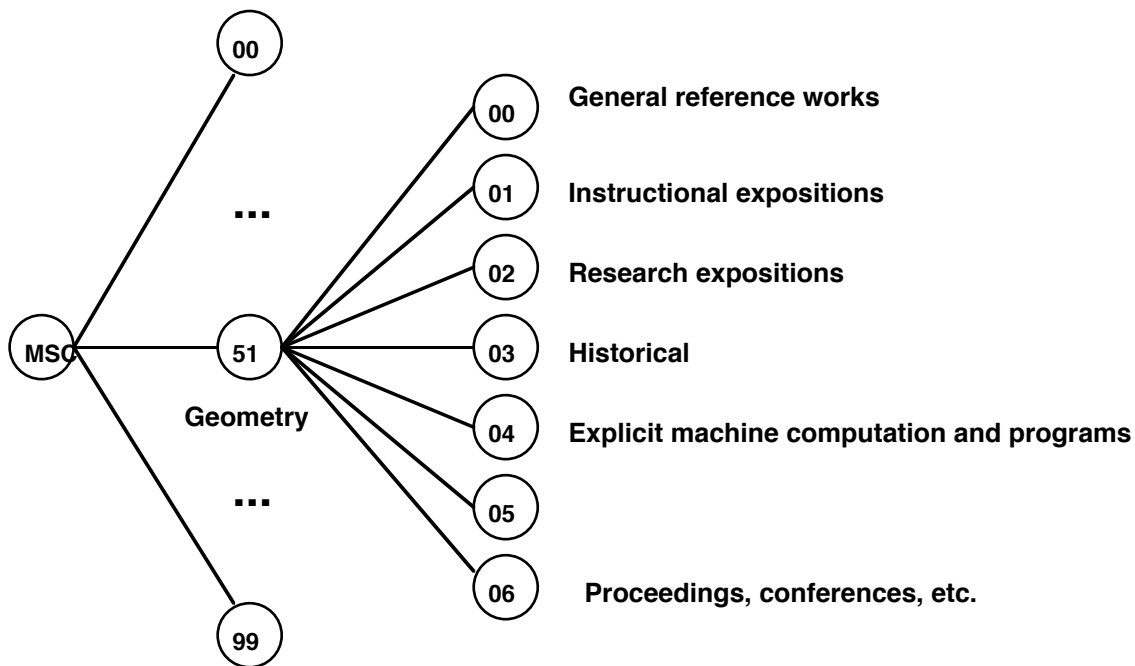Problem: Loss of openness regarding the Internet
- Some openness in combination with WWW only
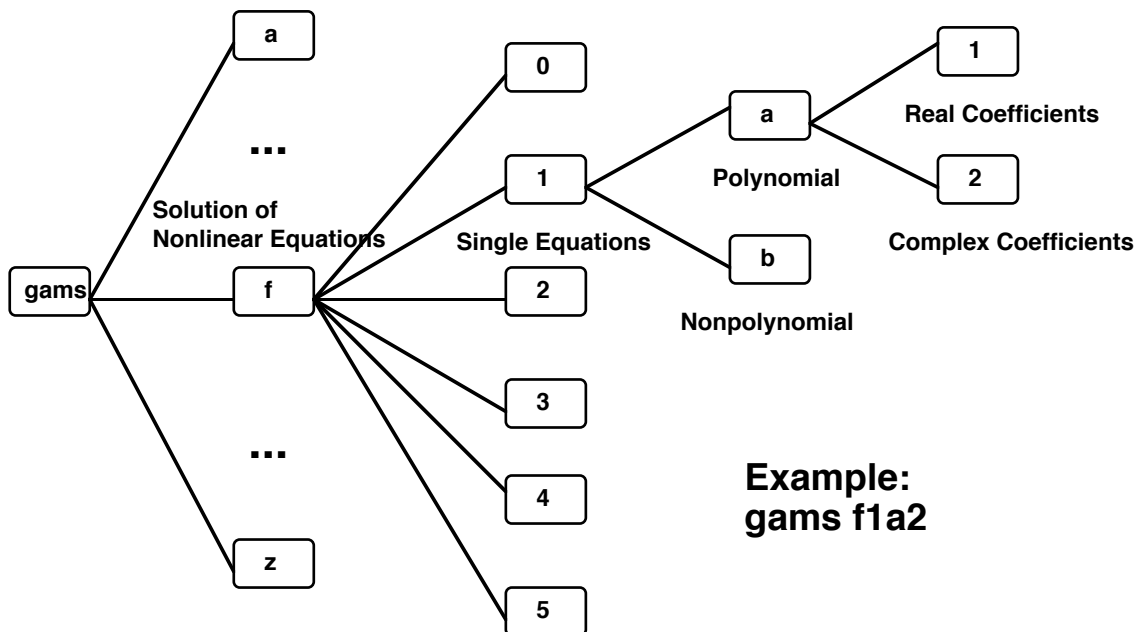- Acceptance by the Internet community?
  Can Hyper-G be distributed over the whole Internet?

Utilization of Classification Schemes

I. MSC - "Mathematics Subject Classification"

- **00** General reference works
- **01** Instructional expositions
- **02** Research expositions
- **03** Historical
- **04** Explicit machine computation and programs
- **05**
- **06** Proceedings, conferences, etc.

51 Geometry

II. GAMS - "Guide to Available Mathematical Software"

Solution of Nonlinear Equations

f — Single Equations

1 — a Polynomial
- 1 Real Coefficients
- 2 Complex Coefficients
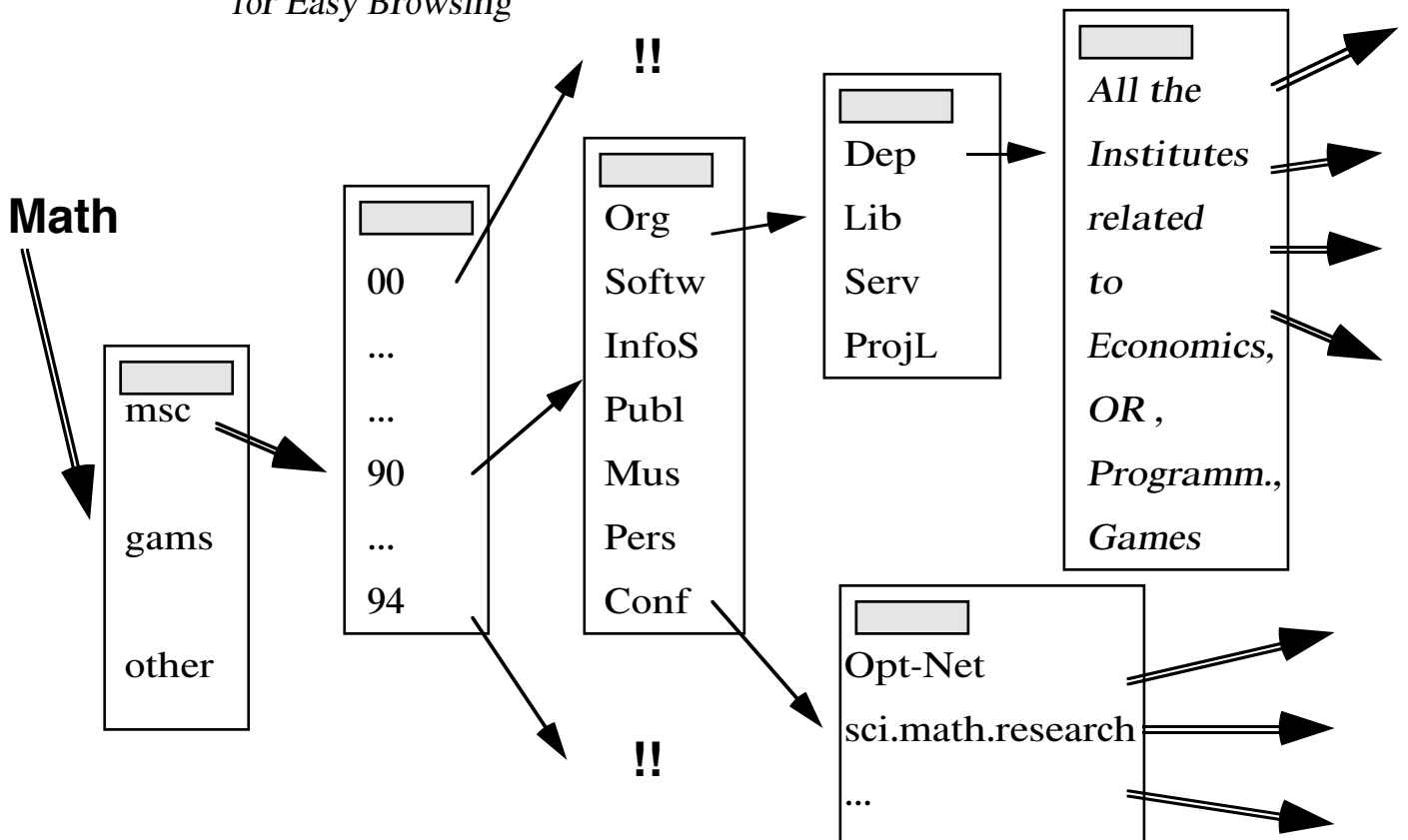
b Nonpolynomial

**Example:
gams f1a2**

III. None

Anything else from the collection structure
will be left unclassified.

Information Retrieval and Distribution

A possible Target
      *Generation of a **Central Hypertext** via a Database*
      for Easy Browsing

!!

Math

msc

gams

other

00

...

...

90

...

94

Org

Softw

InfoS

Publ

Mus

Pers

Conf

Dep

Lib

Serv

ProjL

All the

Institutes

related

to

Economics,

OR ,

Programm.,

Games

!!

Opt-Net

sci.math.research

...

Problems yet to be solved

- •Coordination of local and central maintenance - not possible "by hand"
- • "Wading" through deep nested hypertext structures
- • Simple retrieval mechanism or self explaining query language

But also - for catching "what's new":

- • Avoiding Information "Underload"

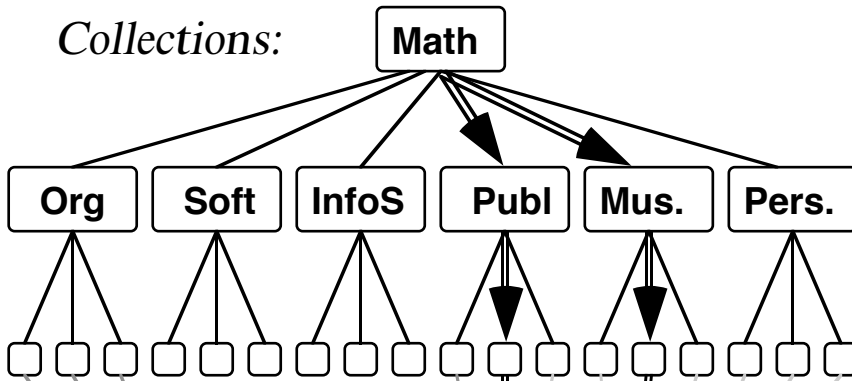*Because it´s not easy, to catch the updates by browsing.*

⇒   **Information Distribution Necessary**
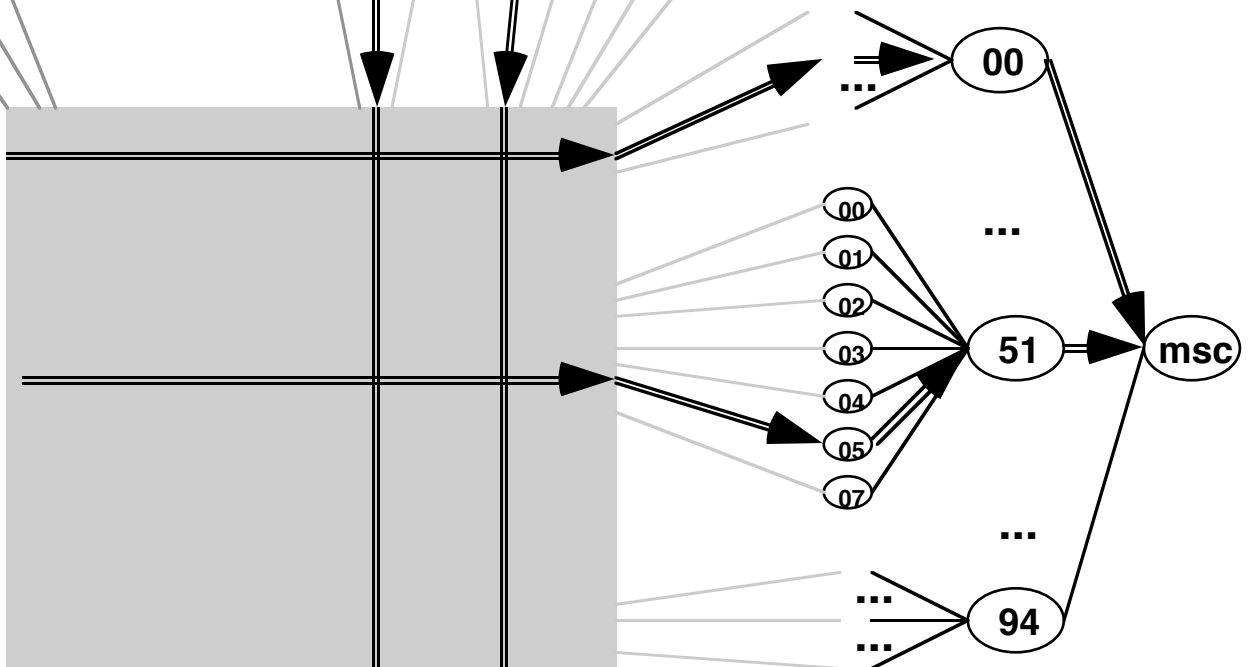
E-mail Based Conferencing

contribution

To: math-net@ ...
Subject: publications, museum
url: http://ftp.zib-berlin.de/pap
classification: msc 00-02, 51-05
content-type: text
*This text is a short description
of a paper at ftp.zib-berlin.de
in a http-location "pap".*

Collections:

Math

Org **Soft** **InfoS** **Publ** **Mus.** **Pers.**

By Message

MSC-Classifications:

... **00**

...

00
01
02
03 **51** **msc**
04
05
07

...

...
94
...

Subscription Examples

Math	anything in the collection hierarchy
Reports	anything "below" reports
Reports, Museum	anything below "Reports" or "Museum"
msc	anything classified according to MSC
msc-51	anything classified "below" MSC 51
msc-51A	anything classified "below" MSC 51A
msc-90, gams-f1a2	anything below both of them ("or")
other	any unclassified item

Any Combination of Subscriptions shall be possible

Information Retrieval & Distribution

1.   Search

a) by Browsing

Hierarchical structures
Web Structures

"what's there" in collection
sorted by subject
linked by title

b) by Query

Database generated Menu/Web-Structures

"what's there" according to query:
• collection items
• msc-index
• gams-index
• keyword in title
• keyword in full text
• author
• date

2.   Receiving

by Message

Electronic Conferencing

"what's new" in collection (or subcollection)
according to MSC- or GAMS-Index
or Other (unclassified) Information

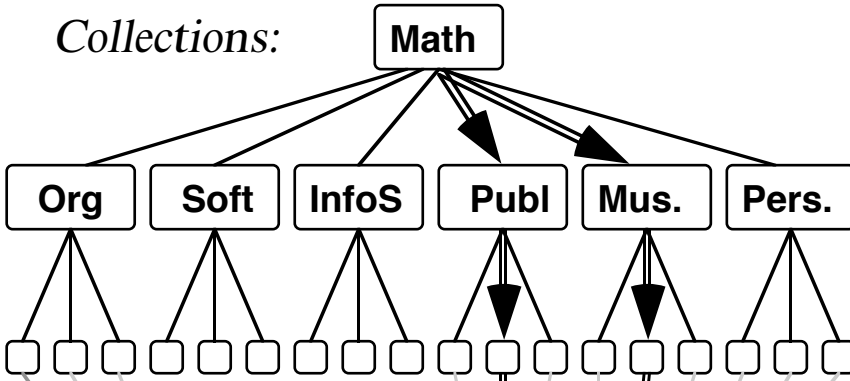Both modes are sensible and necessary:
searching & receiving

Central Math-Net Architecture

Collections:

Math

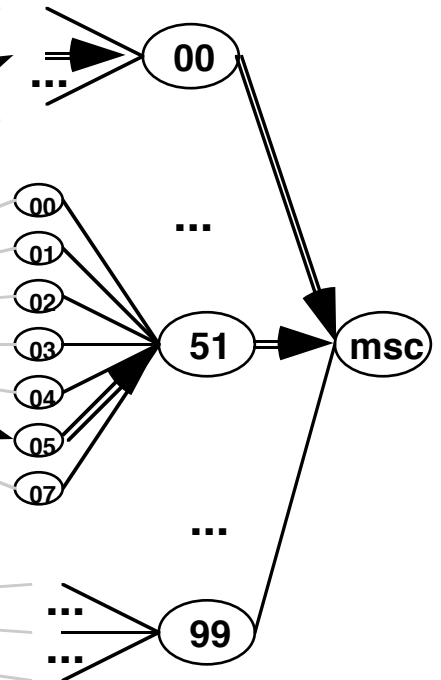**Org** **Soft** **InfoS** **Publ** **Mus.** **Pers.**

contribution

To: math-net@ ...
Subject: publications, museum
url: http://ftp.zib-berlin.de/pap
classification: msc 00-02, 51-05
content-type: text
*This text is a short description
of a paper at ftp.zib-berlin.de
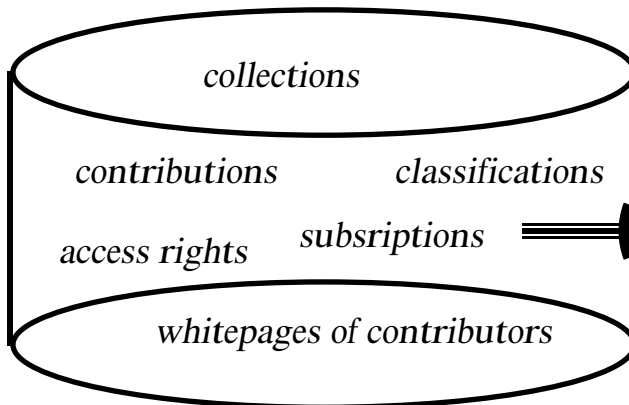in a http-location "pap".*

By Message

MSC Classifications:

00

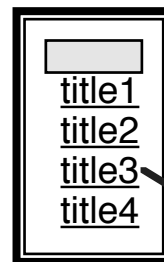...

00
01
02
03 **51** ⟹ **msc**
04
05
07

...

99

...

- central conference system -

By Search/Query

Collection Oriented
Hierarchic Query

collections

contributions   *classifications*

access rights   *subsriptions*

whitepages of contributors

- central database -

title1
title2
title3
title4

*Resulting
WWW-Page*

- local web -

Weak-Structured Mail
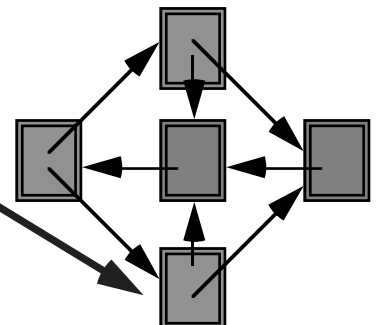
Header:

To:	math-net@...
Subject:	<collection-list> : <title>

Body:

url:	(unified resource locator)
classification:	msc <index-list> ǀ gams
content-type:	text ǀ tex ǀ latex ǀ bibtex ǀ math ǀ mr ǀ postscript ǀ dvi ǀ SGML ǀ HTML library-index

<remaining-body-of-mail> (according to content-type
 used for fulltext indexing only)

Attachm.:

file	(not for redistribution
	will be stored centrally)

implicitely:

address-of-supplier	(derived from header)
nickname	(if registered at math-net)
date	(marks actualization)
title, url, content-type	(if specified via HTML)

Example   *Mail from Mosaic   (sending an HTML-document)*
- positioned at a relevant page referring to art objects

To:	math-net@ ...
Subject:	Museum Modern Art: Anima Computer Art Gallery

*This is sufficient! Anything else will be derived from the mail,
including: title, url, content-type and content (the HTML-text).*

Distributing the Central Part

contribution

**Math-Net
server**

e-mail

B subscribing at A

NNTP

**another Math-Net
server**

A

B

e-mail

NNTP

NNTP to
e-mail gateway

**UseNet News
server**

e-mail

- yet to be evaluated -
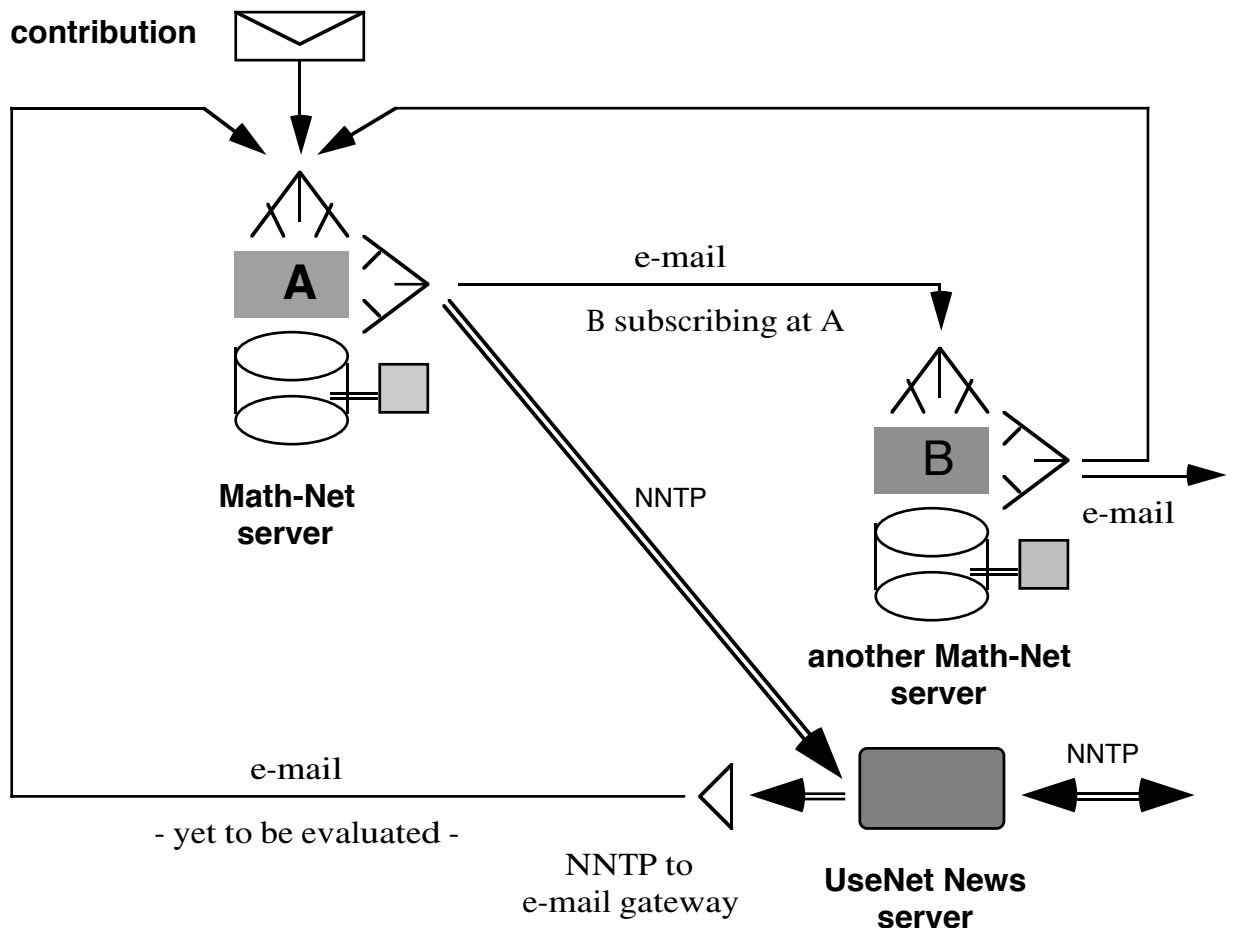
Contributions

- to and from any Math-Net server
- to (and from?) any UseNet News server

Coordination - regarding UseNet News yet to be solved for:
- Administration
- Subscriptions
- Access rights and levels
- ...

Automatic Updates and Recognition of New Contributions

⇒ **Automatic Updates of all WWW Pages and Goper Menus**

- Weekly update cycle over all known URLs
- "Doublettes" are a nonproblem for decentralized information
- Special care for centrally stored information (the attachments only)

⇒ **Automatic Announcement of Updates of Software and Data Collections**
- located in ftp-archives

⇒ **Automatic Announcement of New Preprints, Lecture Notes, ...**
- located in ftp-archives

Assuming some Minimal Standard for ftp-Archives

- Abstracts stored separately
- Naming scheme - such that an abstract can be related to a paper
- Index or Readme file for the whole archive

at the Top-Level of the Archive.

Research Directions

⇒ **Intelligent Agent**
- Mosaic prototype

⇒ **Information Broker**
C1-Projekt of GI "Universeller Informationsvermittler"

Related Work
- Archie
- Indie

Elimination of Manual Work as Far as Possible

Technical Design Principles

Decentralization
- Distributed System Technology
- Minimization of Central Features

Ease of Use
- Hypertext
- Simple Search & Retrieval Techniques
- Electronic Conferencing

Open regarding Internet
- Heterogenity
- International Information Exchange

Consistency
- Local Compatibility

Ease of Maintenance
- Good Authoring Capabilities
- Flexibility

Design for Change
- Extendability
- Adaptability (to new Internet Tools)

Design for Long Term Life Cycles

What will be next?

Discussion Forums and E-Journals

Discussion/Distribution List (UseNet News like)

- merely another collection
- **unmoderated** - contributions by anybody
- **instant distribution** of contributions - which are not stored
- lazy supervision - because of legal questions only
Support for spontaneous and lively discussions

Discussion Forum (e.g. Opt-Net)

- merely another collection
- a **Moderator** is the sole contributor - submissions to him
- he decides on the **value** of the contribution
- and on its **distribution mode** (e.g. weekly digest)
- contents of contributions are stored (text format)
Dedicated quality control by formal responsibility

E-Journal

- merely another collection
- an **Editor** is the sole contributor
- he may be assisted by a **Board**
- distribution of **contents lists and abstracts** (text format)
- **contents** of contributions are **stored** (LaTeX format)
- papers are **printed out and archived** by a library
High Level quality control by peer-reviews

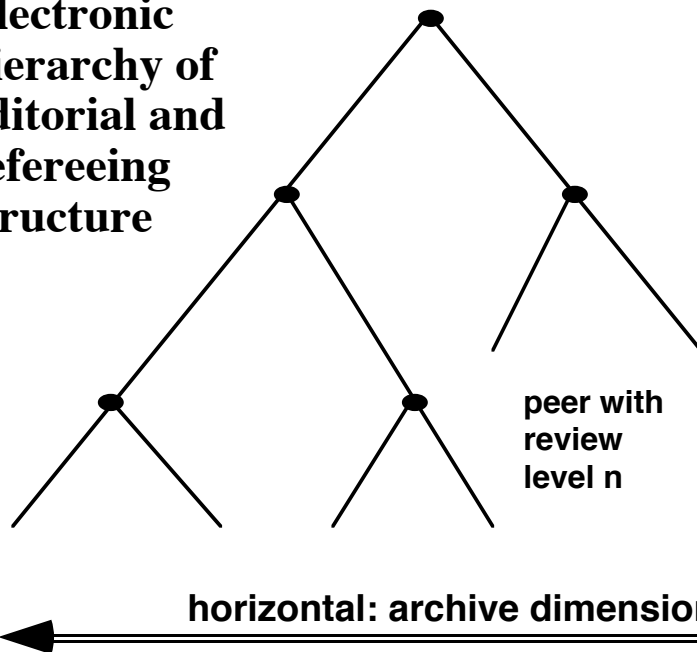**International Cooperation for the Constitution of Boards**

Vision: Electronic Prepublishing

Andrew Odlyzko (AT&T, AMS)
Stevan Harnad (Princeton)

$\Rightarrow$ Scholarly Skywriting

Electronic hierarchy of editorial and refereeing structure

TOP:
top
level
quality

vertical: peer expertise dimension (filter)

bottom
level
quality

peer with review level n

horizontal: archive dimension

Bottom: Global Archive of posted Research Papers
- searchable by classifications, keywords and review level

Research papers as living documents

the related comments, revisions, peer reviews are appended (1 file)

Preprint

Comment

Revision

Peer Review

Nothing submitted can be withdrawn.
Everything public available - via the net.
Notification of arrival to subscribers.

$\rightarrow$ at minimum costs
- technical infrastructure exists

Do we still need
the printed version of
mathematical journals?

Levels of Access Rights

Decentralized Responsibility

All Partners Decide on Their Own Information Offer
- Acceptance and rejection of publications
- Offer of software and data collections
- Availability of organizatorial information (e.g. addresses of staff)

Announcements of Decentralized Information are Permissively Accepted
- Even from anonymous ones
- Validity of URLs will be checked only
Support for Contributions by Individuals

Some - but Mild - Supervision regarding Centrally Stored Information
- Attachments of Contributions
- Selected Discussion Lists
Just to avoid the danger of "illegal" contributions

Permissive Access - Levels of Identification

Unlimited Read Access for Everybody
- Anonymous utilization
- E-mail addresses needed for subscriptions

Semi-Identification for Contributors to Centrally Stored Information
- WhitePage information known to the system
- Nickname and e-mail address known to readers
- Identity verified via signed letter
Anonymous Contributions are Possible in Any Other Case

Full Identification of Coordinators
- Information coordinators of (formal) Partners
- Moderators of discussion lists
- Editors of E-Journals
- Math-Net Administration
Via Veriefied Address and Published WhitePage

Special Care for Certein Collections and Classes
- "MR" and "Zentralblatt" may control MSC-schema
- NIST and ZIB may control GAMS-schema
- Historians may supervise the Ancient Part of the Museum
- Research institutes may supervise the Modern Part of the Museum
- Editors (and their boards) are responsible for E-Journals
- DMV SIGs may moderate their discussion lists

Zentralblatt and Math-Net

Bibliographic References

are the "URLs" of traditional publications on paper, e.g.

Bib-Ref: ACM TOMS, Vol. 12, No. 1, 1994, pp. 69-91

is the universal pointer to the paper

On the Expressive Power of Query Languages
Peter Schäuble and Beate Wüthrich

⇒ Math-Net can also be used for
communications on paper publications

Distribution of
- Announcements, Contents Lists
- Abstracts
- Reviews and Comments

Discussion Forum (Electronic Reviewing)
- Peer-Reviews (closed discussions)
- Call for Peer Reviewing (open)

⇒ **thus gaining new reviewers for new fields of interests**

Publishers - especially smaller ones - may utilize Math-Net
- as Preview Server
- addressing Groups with Special Interests directly
- without the need for maintaining an own server

Zentralblatt may utilize all disseminated information
- Evaluating "quick & dirty" data
- Transforming them into "High Quality" Data

⇒ **Involvement of the Mathematical Community**

Individuals - Small Enterprises

E-mail is the Universal Basis
for Electronic Mathematical Information Exchange

Cooperation of "Isolated" Mathematicians
working or living in some

- Industrial Research Laboratory
- Commercial Company
- Public Institution

- Remote Part of the World
- Eastern Part of Europe
- Developing Country
- ...

should be possible with all - the local and the global - partners of the Math-Net community

Math-Net will be Open regarding the Internet
and may be utilized - either giving or taking - also by individuals

- from schools - teachers or pupils
- from museums (e.g. technical oriented ones)
- interested in history
- scientific or technical writers
- ...

Math-Net may also be Utilized by Small or Medium Companies
- Publishers
- Software houses
- Middle class enterprises with research activities
- ...

wherever they are settled down

Promise of the Internet and the forthcoming "Infobahn"
The Global Village of Virtual Communities