

# Genealogy Network Transfer Protocol

## Note from Conan Albrecht:

**This project will be run very much like an open-source project. All code written will be under the GNU Public License, and the protocol and architecture will be open to all. I will be setting up a CVS tree for source code management in the next few weeks.**

**If you would like to help by writing a client, a directory server, or helping with the protocol, please e-mail me at [conan@warp.byu.edu](mailto:conan@warp.byu.edu)**

Brigham Young University  
Family History Conference

March 29, 2001

Marriott School of  
Management  
E-Business Center

# **Genealogy Network Transfer Protocol**

**Conan C. Albrecht, Douglas Dean,  
Robert B. Jackson, Stephen W.  
Liddle, Rayman D. Meservy**

**Brigham Young University  
Family History Conference**

**March 29, 2001**

**Marriott School of  
Management  
E-Business Center**

# Current Limitations

## Non-Digital Formats

- Not searchable in automated ways
- Single, unique books exist in different libraries around the world

# Current Limitations

## Central Repository (LDS Church)

- Voluminous
  - ◆ Hard to index and search
  - ◆ Costly to store and access
  - ◆ Many delays to publication
- Management (**most complex problem**)
  - ◆ Resolution of data conflicts
  - ◆ Owners of data are #1-qualified caretakers
- Multiplicity of central repositories

# Current Limitations

## WWW

- Keeps control with owners, but
- Unstructured data formats
  - ◆ HTML is a presentation language, not a structured data type language
  - ◆ Hard to search in automated ways
  - ◆ Requires very intelligent agent software
  - ◆ Page scraping

# Potential Solutions

## WWW

1. Create smarter agents that can scrape page with near-100% accuracy
2. Make the data more structured

# Potential Solutions

## WWW

1. Create smarter agents that can scrape page with near-100% accuracy
2. **Make the data more structured**

# Current Limitations

## XML

- Distributed and structured, but
- A document type definition only structures the data
- It does not define how the data is accessed, where it is located, what port to access it through, etc.
- What automated sharing benefits does it give us beyond GEDCOM?



# Questions That Must Be Answered

- **Why** will I share my data?
  - ♦ Genealogical researchers are eager to share
- **What** data will I share?
  - ♦ Structured genealogical data: family group records, individual records, etc., in XML or GEDCOM format
- **Where** will I share it with them?
  - ♦ The Internet is a ubiquitous network
- **Who** will I share data with?
  - ♦ A **registry** or **index** is needed
- **How** will I share it with them?
  - ♦ A **common transfer protocol** must be defined
  - ♦ A **common searching language** must be defined

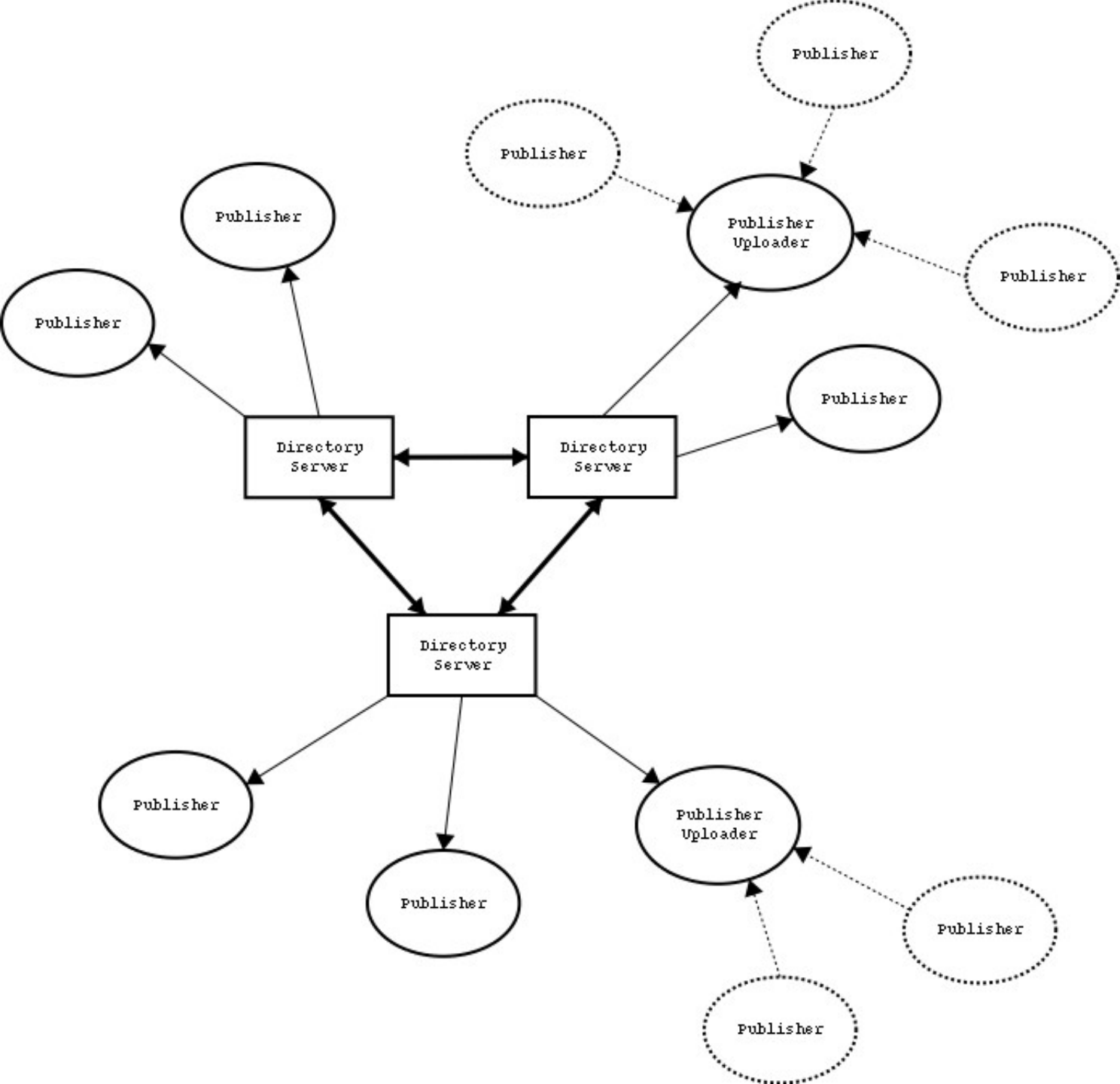
# GNTP

- Protocol and architecture  
(not an application)
- Peer-to-peer
- Central ring of directory servers
- Outer ring of data providers (individuals)

# Individual Node Types

- Directory Server (indexes publishers)
- Publisher (publishes source data)
- Publisher/Uploader (accepts uploads from non-technical or non-connected researchers)
- Searcher (simply searches the network)

# GNTTP Architecture



# GNTTP Protocol

- Defined similar to SMTP, HTTP, POP3, and other standard protocols
- Includes commands for:
  - ◆ Publication on the network
  - ◆ Directory server indexing of nodes
  - ◆ Node-to-node searching
- Standard directory server and inter-node searching language

# GNTP Benefits

- Supports exact searching with structured data types
- Independence from WWW reduces potential for illicit information on network
- Retains control with users (little overhead to data corrections, additions, etc.)

# GNTP Benefits

- Allows immediate, real-time updating
- Simple clients with automatic publication and searching allow less-advanced users to participate

# GNTP Benefits

- Distributed as free software (GPL) and protocol definitions, opening boundaries between organizations
- We expect many organizations will open their databases as nodes on the network
- Organizations can still maintain revenue streams as uploaders for researchers



# Status

- GNTP 1.0 protocol is nearly finished
- Reference implementation client and initial directory server will be finished by June
- Web site and directory server go live July 1, 2001
- LDS Church will create a link on [familysearch.org](http://familysearch.org) to the “experimental” network
- LDS Church will provide genealogical database for initial publisher node

# Involvement

- Research into the most efficient architecture for the inner-ring of directory servers is needed
- Node interface to the Church genealogical database needs to be developed
- Several end-user, publisher clients need to be created (Windows, Mac, Unix, Java, etc.)