

Structure of a historical data archive

Bálint Dömölki

John von Neumann Computer Society

IT History Forum

An early publication related to *Digital Humanities*



Kalmár László - Telegdi Zsigmond (szerk.): ÁLTALANOS NYELVÉSZETI TANULMÁNYOK 2. (Budapest, 1964)

Dömölki Bálint—Fónagy Iván—Szende Tamás

Köznyelvi hangstatisztikai vizsgálatok*

*Dömölki Bálint munkája a 3. fejezet; Fónagy Iván az 1., a 4. és az 5. fejezet és a 2. fejezet 2.1., 2.2. pontja Szende Tamásé a 2. fejezet többi pontja;

First „non-numerical” computer application in Hungary (1961)

Two statements

- I. „Technological heritage” is part of Cultural Heritage
=> a data archive for the history of computing in
Hungary may be regarded as a Digital Humanities
activity

- II. This archive can be generalized into an abstract model,
covering a wide range of cultural heritage archives

European Year of Cultural Heritage

Cultural heritage comes in many shapes and forms.

- **tangible** – for example buildings, monuments, artefacts, clothing, artwork, books, machines, historic towns, archaeological sites.
- **intangible** – practices, representations, expressions, knowledge, skills - and the associated instruments, objects and cultural spaces - that people value. This includes language and oral traditions, performing arts, social practices and traditional craftsmanship.
- **natural** – landscapes, flora and fauna.
- **digital** – resources that were created in digital form (for example digital art or animation) or that have been digitalised as a way to preserve them (including text, images, video, records).



Technology heritage \subset Cultural heritage

Collection and distribution of factual data about the history of different fields of technology can be regarded as an important part of the preservation of our cultural heritage

- „knowledge, skills - and the associated instruments” may include technological knowledge etc.
- „traditional craftsmanship” is time-relative:



- Data archives can be built to collect data about the history of a specific technology field in a given geographical area and time period.

How to preserve?

- In the real world (physical preservation of „artefacts“):
 - Museums, exhibitions ...
 - Memorial meetings,
 - Environment protection
 -
- In the virtual world (preservation of information about...):
 - Catalogs,...
 - Virtual museums, exhibitions, ...
 - ...

= Archives



Museum of Fine Arts, Budapest



The Web Gallery of Art is a virtual museum and searchable database of European fine arts from the 8th to 19th centuries.

Basic components of a generic model for archives

- *Domain* is a restricted part of the real world we are interested in.
- *Items* are elements of the domain, having a predefined set of *properties*
- Value of a *property* can be either a simple data (number, text, etc.) or a reference to another item.
- *Basic items* are associated with a file stored in a *repository*
- *Composite items* are defined as virtual entities, not associated with a single file in the repository, but – through their *properties* – being connected to other – simple or composite - items

*European fine arts from
8th to 19th centuries*

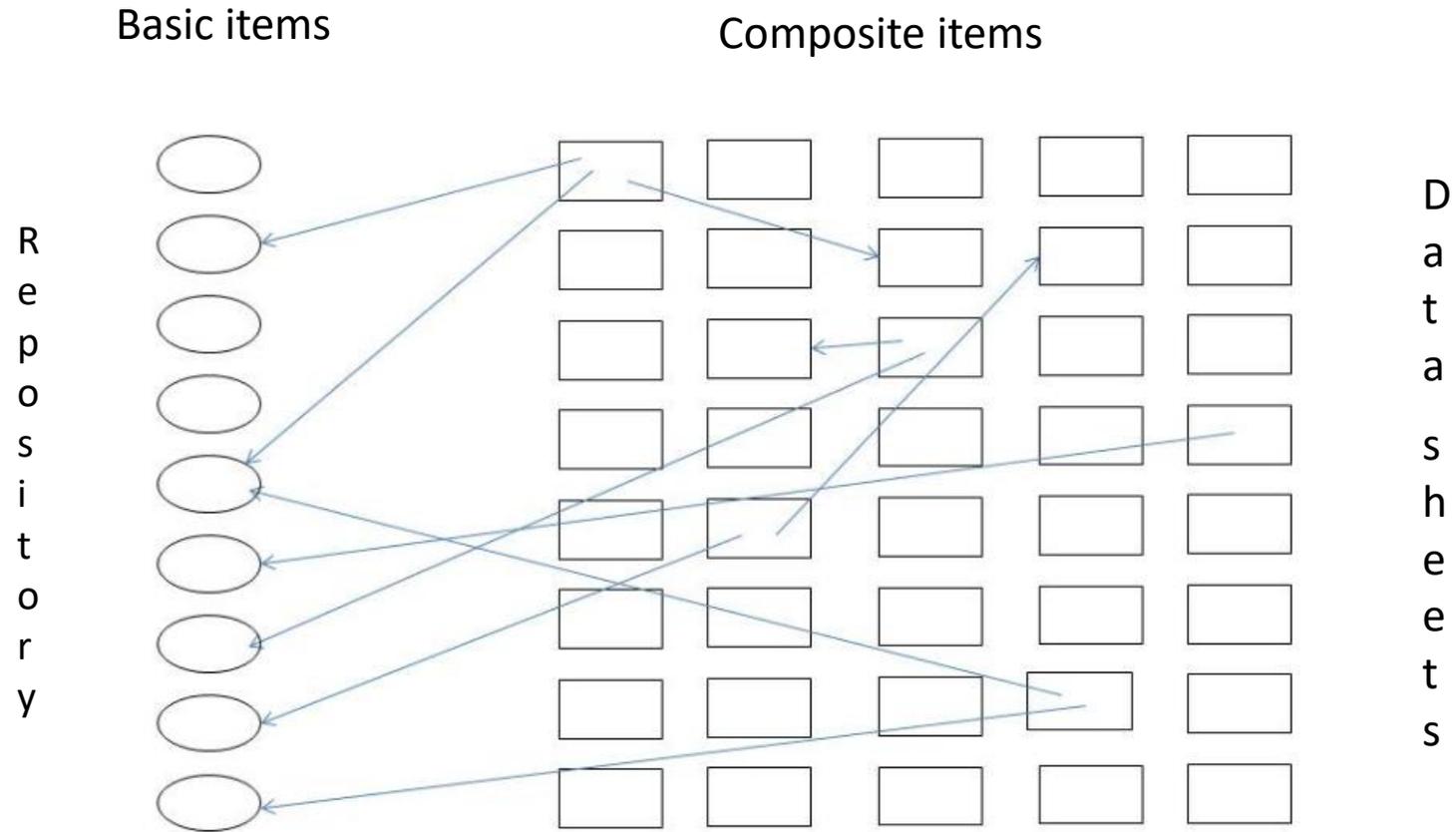
*(images of) Artefacts
(paintings, sculptures, ...)*

*Artists, museums,
artistic schools , ...*

Structure and tools of the generic model

- *Items* with the same *property* structure are organized into *sections*
- Items in each section can be viewed in the form of *data-sheets*, presenting all information about the item and also in the form of *tables*, listing one item per row, with the most important properties;
- For each *section* there exists
 - a) A *description* of its property structure, defining the *fields* of the *datasheet*,
and the following tools
 - b) An *editor*, allowing to input all the fields of the datasheets
 - c) *Screenprinter(s)* to produce the appearance of the datasheet on the screen
 - d) *Table-builder(s)*, producing a table of the section, with all datasheets as rows and values of selected fields in the columns

Data structure



Additional features of the generic model

- Allowing the widespread use of *cross-referencing* within the archive and also to objects of the outside world;
- Including a sophisticated *search system* for finding individual items or lists of items with common properties;
- Providing tools for creating and *editing* the archive, with the possibility of parallel building of (inter-related) sections;
- Organizing *interaction* with the users, allowing them to send feedbacks, including proposals for modification/addition.
(= *limited crowdsourcing*)

Hungarian IT History Archive

- The generic model is demonstrated on the example of a data archive for the history of the Information Technology field in Hungary
- The Archive is created by the *IT History Forum* (iTF) of the *John von Neumann Computer Society* (NJSZT)
- **Mission** of iTF: to collect and preserve memories, while available...
 - memorial conferences, oral history video interviews
- The Archive is being developed by a team of 15-20 volunteers



„To protect the values of the past, to adapt to the present, to influence the future.”



IT history collections in the world

Germany



Well organized overall exhibition with detailed information about the exhibits published on-line

USA



General exhibition plus several partial shows on selected topics, some of them on-line only, in the form of a virtual exhibition,



Extensive repository of documents and oral history interviews, plus research papers on different topics.

Russia



Very detailed on-line collection of documents and papers about different aspects of domestic computing (players, results, etc.), many of them available in English translation

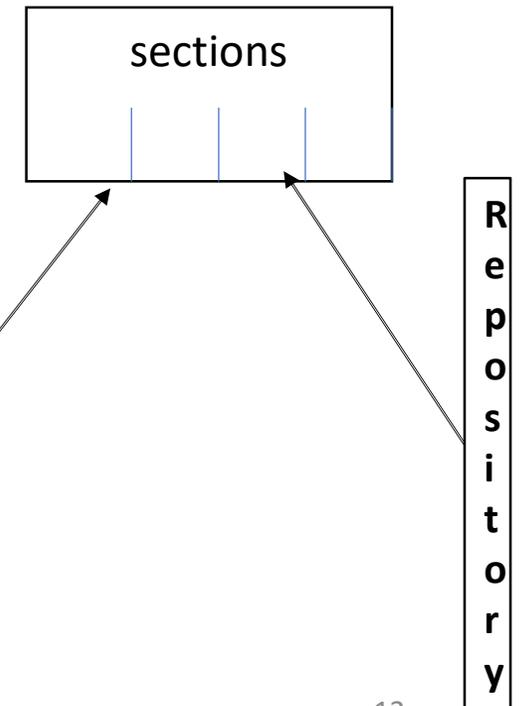
Our approach

is a mixture:

- Based on an on-line repository of documents, papers ,videos etc. being collected.
- Building upon it a *superstructure* of organized collections of data about players , results and events of Hungarian IT history
- Working in a close cooperation with the *IT History Exhibition* (ITK) of the NJSZT, located in Szeged, collecting the physical objects related to the history of IT in Hungary



Large collection of computers used in Hungary in the past 60 years. Only collection containing both Western and Eastern computers



Instantiation of the generic model

The *domain* is the history of the Information Technology field in Hungary, from the 50s to present days, with emphasis on the years before 1990.

The *basic items*, being collected are

- *Writings* (books, publications, manuscripts etc.) related to Hungarian IT history,
- *Recordings*, including oral history interviews, videorecording of talks at iTF events, historic photos etc.,

Composite items include

- *Persons*, having some relation to IT (both alive and deceased),
- *Institutions* (in business, research, education, administration etc.) having been active in IT,
- *Products* (hardware, software) developed and/or intensively used in Hungary,
- *Events* (including conferences, seminars, workshops, contests etc.) organized in Hungary in the past 60 years.

Implementation issues

- The Archive can be reached on the homepage of the IT History Forum at itf2.njszt.hu (presently in Hungarian, English version is being prepared).
- Implementation is done in WordPress. In some extent we are making use of some WordPress specific solutions, the basic structure and algorithms, however, can be adapted to other software environments
- A few specific challenges
 - Due to the parallel development of the sections it is possible to refer to an item not existing yet in that section
 - Names of the institutions may change time to time, but we have to use everywhere the "historical" names.
 - Personal data protection considerations should be taken seriously. (GRDP!)



Neumann János Számítógép-tudományi Társaság

INFORMATIKATÖRTÉNETI FÓRUM



[NEWS](#)

[ABOUT US](#)

[CONTACT](#)

[ITF EVENTS](#)

[ARCHIVE](#)



[PERSONS](#)

[INSTITUTIONS](#)

[PRODUCTS](#)

[CONFERENCES](#)

[WRITINGS](#)

[RECORDINGS](#)



May 30, 2018

DH_2018 Budapest

16

Raffai Mária	Computing Behind the Iron Curtain and Beyond	iTA 58	2006					
Álló Géza Képes Gábor	A jövő múltja	iTA/665	2013	nincs	ITK			
Malinovsky, Boris Nikolaevich	Pioneers of Soviet Computing	iTA/462	2010					
Dömölki Bálint Sántáné-Tóth Edit	Formal Descripton of Software Components by Structured Abstract Models	iTA/673	1976		magán			 
Varga Ákos Endre	The KFKI TPA series: much more than just “clones”...	iTA/482						
Sántáné-Tóth Edit	Computer Oriented Higher Education in Hungary – the Beginnings	iTA/910	2017		magán			
	IT STAR papers on the history of computing	iTA/874	2018					
Dömölki Bálint	The story of the first computer in Hungary	iTA/875	2018		magán			
Dömölki Bálint	Computing in Hungary – Through the History of Five Institutions	iTA/872	2014		magán			
Szabó Máté	Hungary’s Early Years in the Ryad	iTA/855	2015					
Szabó Máté	The M-3 in Budapest and Szeged	iTA/855	2016					

Co-operation possibilities

- *The structure of the Archive and its software solutions might be adapted to the creation of similar historic data archives (or their parts) for other subject areas.*
- The generic model and its implementation in the IT History Archive can be used to help similar activities for other domains on different levels:
 1. To invite items (e.g. persons, products etc.) from a different subject area into our Archive
 2. Create subsections for the items from the other domains and making these subsections accessible separately
 3. Create new sections, specific to the other domain
 4. Share the know-how of creating our archive with the developers of the archive for the other domain
 5. Start a joint project for developing and/or using an archive for both (or more) domains

Thank you for your attention