

What's Next?



Data sharing, Attribution, and Rethinking CIS

Rebecca Snyder, NMNH | snyderr@si.edu

Let's Go Back...

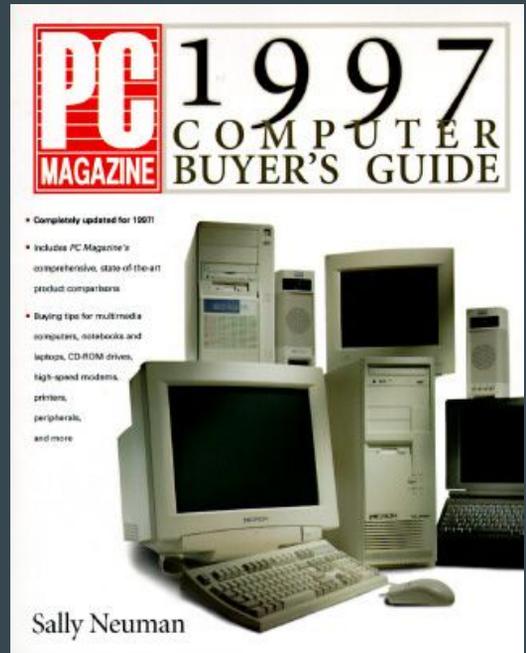
~1997

Yes, that's Times New Roman you font nerds. <3

Why 1997? That is when a large bulk of the initial design of the TexPress backend was created. Test clients out in 1998, first prod releases in 1999.

1997...

Your computer looked something like this:



So much beige...

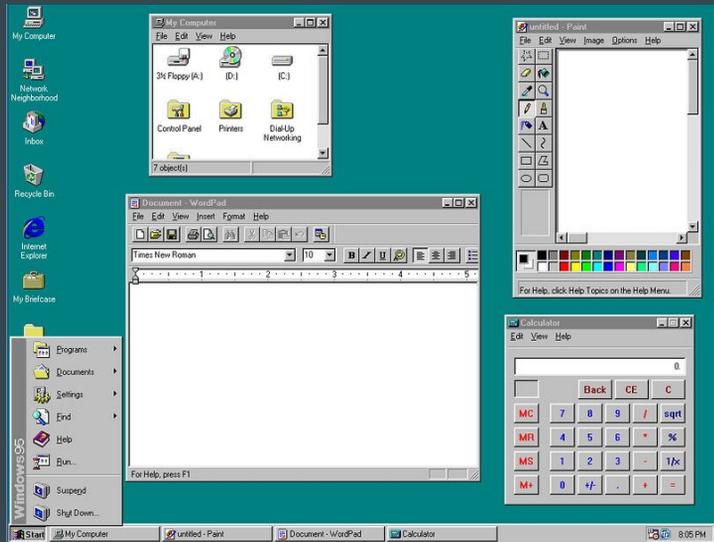
1997...

Windows 95 was the new “modern” OS:



1997...

We were excited to
be free of Win 3.1!



1997...

This was probably your internet browser:



IE v 4.0 was a very distant 2nd in popularity.

1997...

This was your search engine:



No Google till 1998, and then not widely adopted for years.

1997...

Or if you were cool:

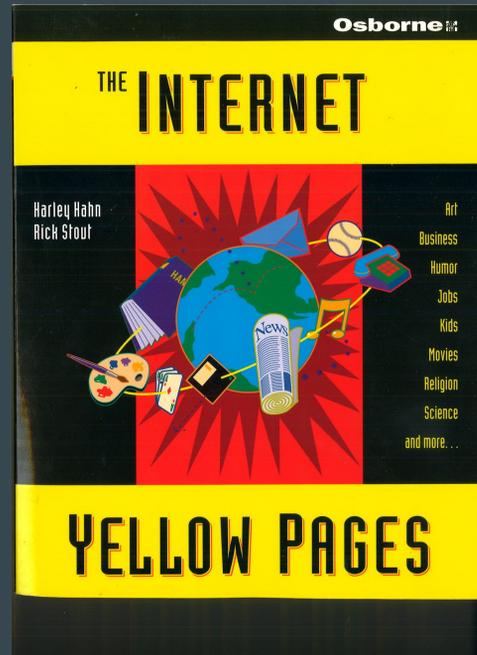


1997...

You could go to 1.1m websites! Much better than last year's ~250,000!

There were so many options, they made a "phonebook"! All 447 pages!

NMNH early databases and gopher sites are listed even!



1997...

...and with your Win 95 box, your AltaVista browser, and your brand new modem, you could visit all those websites at a whopping 56kbps! Twice as fast as last year's 28.8kbps.

**Today's 1gbps speeds are over
150,000 times faster.**



56kb = 0.0000065gb

8589934.6kb = 1gb

1gbps is 153,391.7 times faster than 56k (in theory)

Like in the 56k days, no one ever gets max capacity 1gb.

1997 State of Standards

Dublin Core was only an idea 1995-1998 (and not an ISO until 2007)

The very first baby version of DarwinCore (the NSF ZBIG version) wasn't until 1998

The *long-lived* "draft standard" of DwC wasn't until 2007

The XML DwC schema wasn't even until 2009

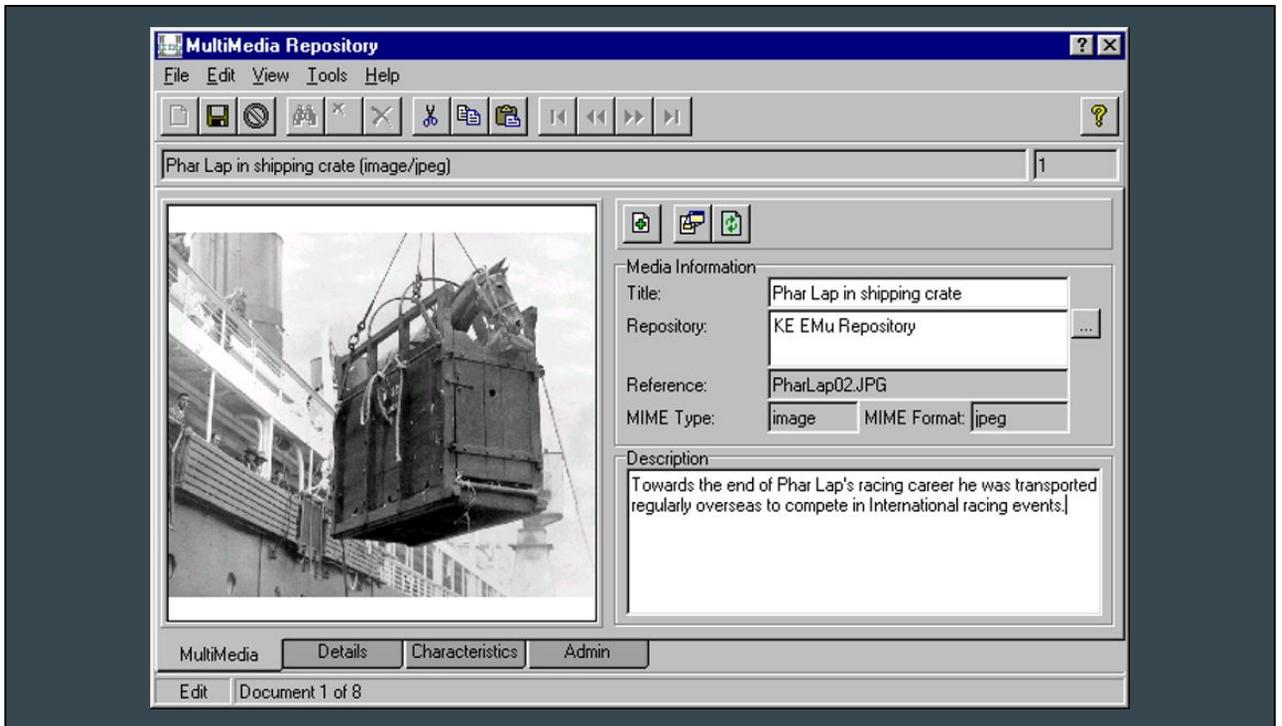
Why does this matter?

This is the environment the first EMu client was created in, leading up to the first official full release in 1999.

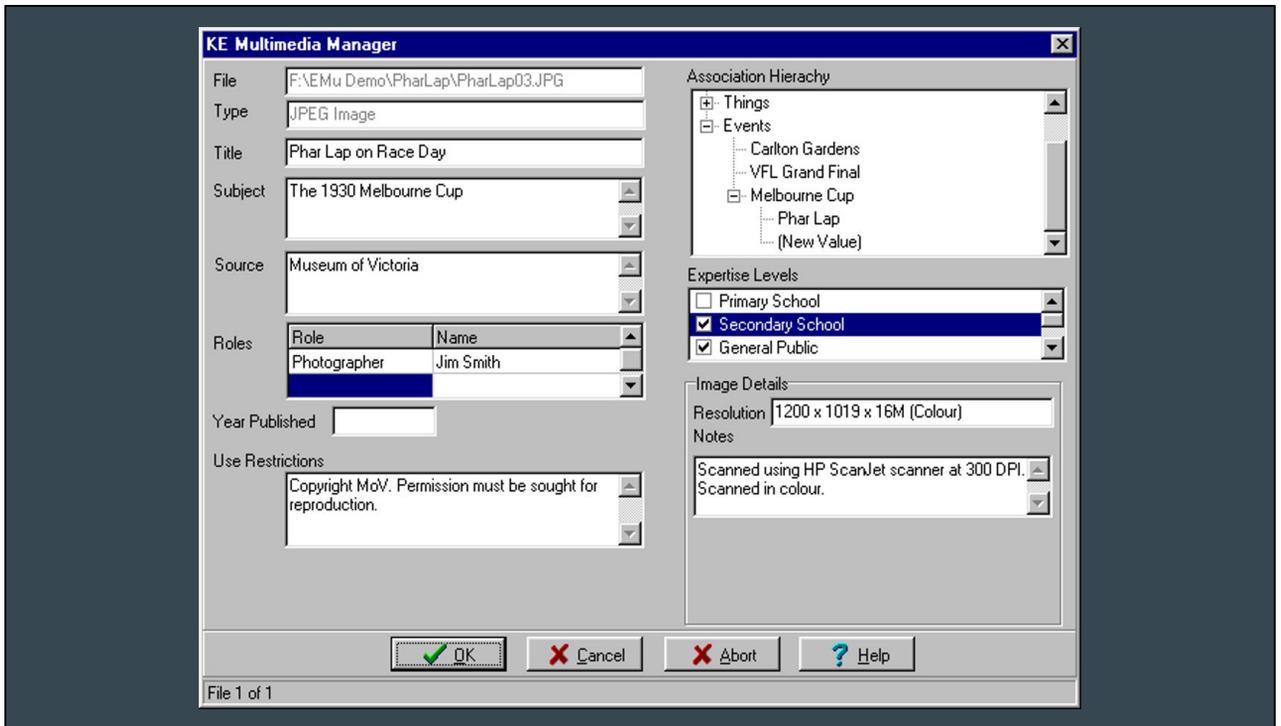
The database had to accomplish all the functionality museums needed, without being able to look outside itself for information or share with others.



Client screenshots circa 1997-1998 and graciously provided by Ian Turnbull.

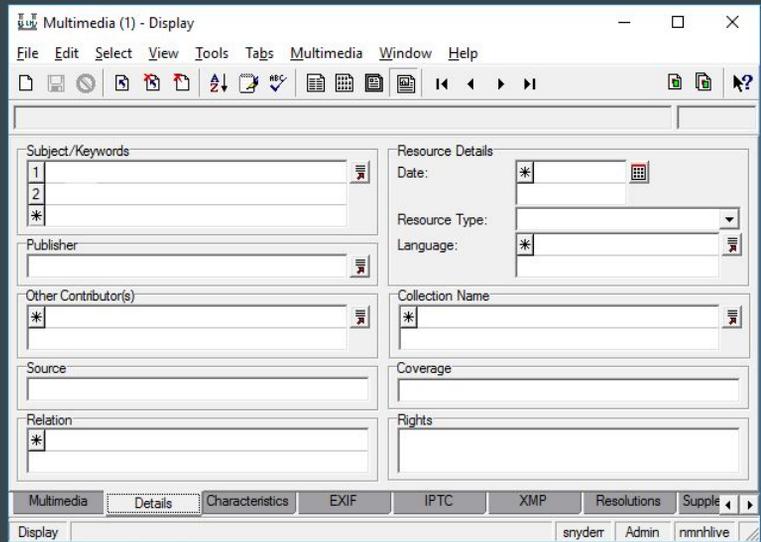


The look and feel is amazingly familiar!



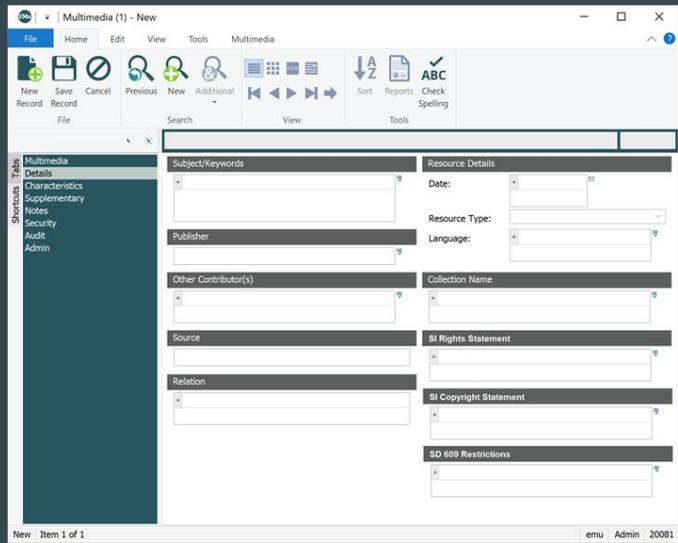
I remember these examples in our training clients!

2000's



Client "just got a little work done" (light facelift)

2019/2020



Major facelift: but it's still the same underlying backend structure.

What can we do now that we couldn't before?

Let's leverage modern technology, updates to computer science and database theory to do more with less.

Let's allow our collections managers to focus on the data specifically about their specimens and outsource the periphery.

Link to trusted data resources for things like localities, known parties, bibliography, taxonomy (wouldn't it be nice to only think about "filed-as" and leave "currently accepted" and synonyms to someone else???)

All those local copy fields... with CDP we don't need them anymore!

Data Sharing and Attribution

This can happen in a variety of ways.

With the CDP switch, we can leverage external indexes.

- EX: ITIS for taxonomy. Either use it to just expand search, or populate data
- What else could we reference? LOC terms? VIAF? ORCHID?

We could link to external data sources - with content updating live

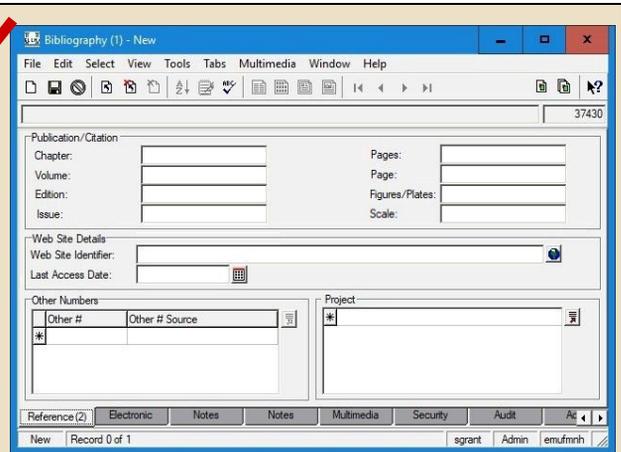
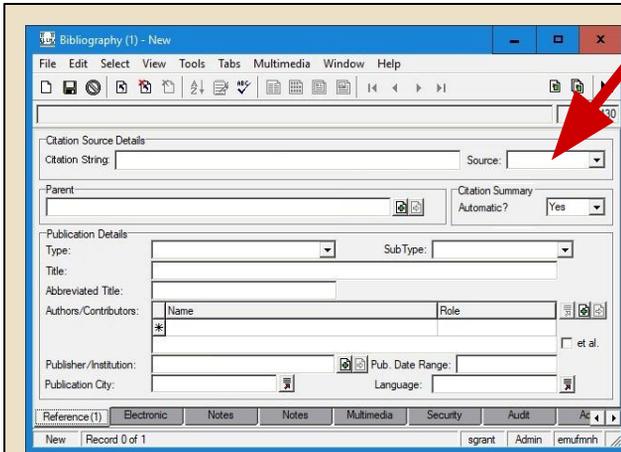
- Linked data in an external access layer - with or without database fields
- Store attribution/data-source
- EX: Material samples data, taxonomy, what else??



Validate Reference Tool

Purpose: This tool uses web services to retrieve data from a selected online bibliographic resource [Source] and populates the record with returned data.

- Pass the value entered in CitCitingText or DOI (can only enter one or the other) to the service selected in NEW Source field.
 - **Parent** : *JouParentRef* - This value is prefilled from the underlying record if the record has been saved.
 - **Citation String** : *CitCitingText* - This value is prefilled from the underlying record if the record has been saved.
 - **DOI** : Has to be pasted in manually
 - **Source** : *NEW Field* - This value is prefilled from the underlying record if the record has been saved.
 - 1) BHL (<https://www.biodiversitylibrary.org/>)
 - API
<https://about.biodiversitylibrary.org/tools-and-services/developer-and-data-tools/#APIs>
 - Data Elements
<https://www.biodiversitylibrary.org/docs/api3.html#elements>



Citation String = user supplied citation
 Source = WorldCat; BHL
 Automatic? = yes/no (executes the Validate References tool)
 Project = for web development and tagging projects
 Other Numbers = Not new, but redesigned to handle multiple number types.

The rest of the fields currently exist. We've eliminated the duplicates and have chosen a single instance of that field type. All "Type" records use the same backend fields.

We are eliminating all but 2 tabs from the bibliography module. The fields in each tab can apply to any "type" of bibliography record, the the type of record is determined by the parent it is linked to.

RS: red arrow pointing to attribution element. "I got this data here. If wrong, blame them. But i still can search on it"

What do we want to do?



Open discussion. Let's "talk amongst ourselves".

Gifs: 2001 & Mike Myers. Fair use, no copyright implied.

System-wide Functionality

Wouldn't it be nice...?



Sharon Grant

Janeen Jones
Kate Webbink
Pete Herbst
Rob Zschernitz



North American
User Conference
2019

Intro

...if my database would help me find things?



Things that would help me/us/you.



- Query methods
 - Count rows in multi-valued tables
 - Count same/different values in tables
- Sort default for queries through attached fields
- ExtendedData
 - Default Sort of Records during Data Entry
- OpenRefine like thingies native to the client
- More embedded collaborative tools - community to tool building, example Specify, github, wouldn't it be nice to have an EMu Github community that had common formats for documentation and toolsets

[The only way out lies within your mind.]

https://docs.google.com/document/d/1LLR-5jqcsQF4qqtTBUH6EkG0EvmLBuYmTCs_dvk5JsKI/edit#heading=h.adtb2dl58rfh

...if I could fix it before it breaks?



Things that would help me/us/you. 

- **Development of Help functionality**
 - Tools Help
 - Tab Help

- **Admin controlled messaging**
 - Parent Attachment Warning

[The only way out lies within your mind.]

https://docs.google.com/document/d/1LLR-5jqcsQF4qqtTBUH6EkG0EvmLBUYmTCs_dvk5JsKI/edit#heading=h.15o1txt5jg5n

...if I could fix things after they break?



Things that would help me/us/you. 

- Recall
 - Make it work

[The only way out lies within your mind.]

https://docs.google.com/document/d/1LLR-5jqcsQF4qqtTBUH6EkG0EvmLBUymTCs_dvk5JsKI/edit#heading=h.vfqf8bcar4jq

...if reporting was better?



Things that would help me/us/you. 

- Pipe-delimited table-values report-option

[The only way out lies within your mind.]

https://docs.google.com/document/d/1LLR-5jqcsQF4qqtTBUH6EkG0EvmLBuYmTCs_dvk5JsKI/edit#heading=h.yup7u5a56tfk

...if fields worked more intuitively?



Things that would help me/us/you.



- Weighted ordering of hierarchy
- Multimedia tab detail list view

[The only way out lies within your mind.]

https://docs.google.com/document/d/1LLR-5jqcsQF4qqtTBUH6EkG0EvmLBuYmTCs_dvk5JsKI/edit#heading=h.adrd2cvpe7bl

...if the registry sucked less?



Things that would help me/us/you. 

- Registry entry to create conditional mandatory based on more than one field
- Remove functionality that causes whole group box of fields to be turned off if the last field in the group box is turned off
- Give us a registry option to make whole tabs read-only for user groups

[The only way out lies within your mind.]

https://docs.google.com/document/d/1LLR-5jqcsQF4qqtTBUH6EkG0EvmLBuYmTCs_dvk5JsKI/edit#heading=h.5gy97vpqkmhc

Questions, Comments, Thought?

Yes, it would.



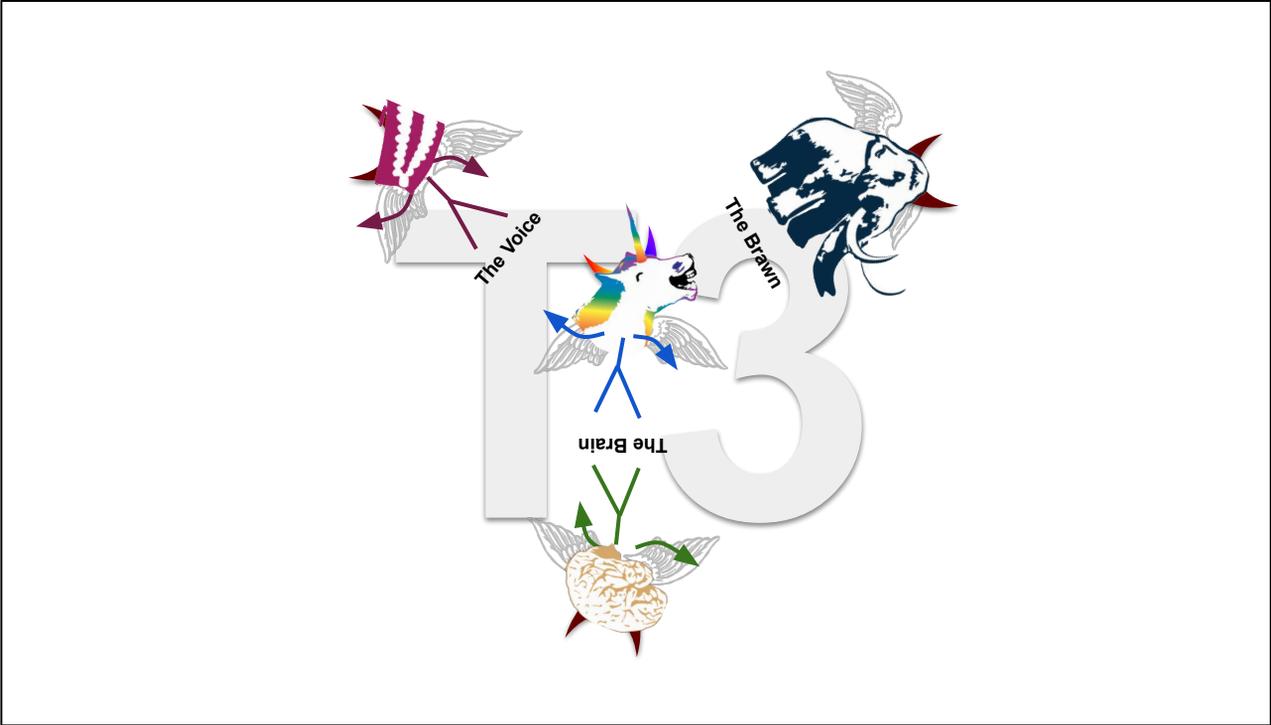
Sharon Grant

Janeen Jones
Kate Webbink
Pete Herbst
Rob Zschernitz



**North American
User Conference**
2019

Acknowledgements



Created by Kate Webbink