# Economic Growth Center Digital Library a project funded by the Andrew W. Mellon Foundation

Ann Green, Steve Citron-Pousty, and Marcia Ford Social Science Research Services Yale ITS Academic Media & Technology

Sandy Peterson and Julie Linden
Yale Social Science Libraries and Information Services

Christopher Udry
Professor, Yale Dept of Economics
Director, Economic Growth Center and Dept. of Economics

## Goals of the grant: Improving access to statistical resources not born digital

Build a prototype archive of statistics not born digital.

Implement standard digitization practices and emerging metadata standards for statistical tables.

Document the costs and processes of creating a statistical digital library from print.

Build the collection based upon long range digital life cycle requirements.

Present the prototype digital library to the scholarly community for evaluation.

### **Research questions**

- ## What effect does online access to the statistical information have on scholarly use of the materials?
- ## Are common digitization practices and standards suited to statistically-intensive documents?
- **\*\*** What are the long term preservation requirements of the EGCDL?
- ## What are the costs of producing high quality statistical tables with OCR and editing?
- ## How scalable is this process, for what kinds of collections, and for what purposes?

## **Digitization process**

- Vendor review:
  - Part one: request for proposals (13 vendors)
  - Part two: production of samples and extended proposals (6 vendors)
  - Part three: final review and budget evaluation
  - Part four: contract negotiation and processing
- Prepared and shipped 221 volumes in Fall '03
- All materials received back at Yale Feb '04
- Quality assessment period complete March '04

## Deliverables received: images and PDFs

- 300 dpi TIFFs of each page of each volume, including cover and back of volume (102,534 TIFF files; 415 gigabytes)
   Color TIFFs for color pages, black and white TIFFs for
- pages with only black and white
   PDF image + text files of each chapter of each
- PDF image + text files of each chapter of each volume (5,607)
  - Separate files for each subject chapter, front matter, indices, and back matter

## Deliverables received: statistical tables in Excel

- Excel tables (16,488) of demographic and economic tables for 1994, 1996, 1998, 2000
- Selected OCR'd statistical tables converted to Excel format
- DSI operators reviewed each table twice
- Custom tagging done by DSI improves ability to extract columns and rows into online database

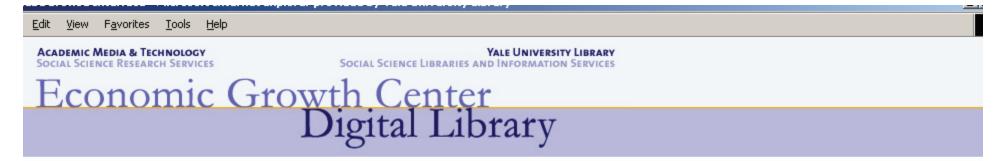
## Automated metadata production for PDF and Excel files

- Defined metadata format for PDF and Excel documents
  - Dublin Core subset (title, date, format, identifier, source, coverage, subject)
- Wrote scripts to produce individual metadata records for each PDF chapter and each Excel file

## First generation interface: Select, view, download ssrs.yale.edu/egcdl

### Features:

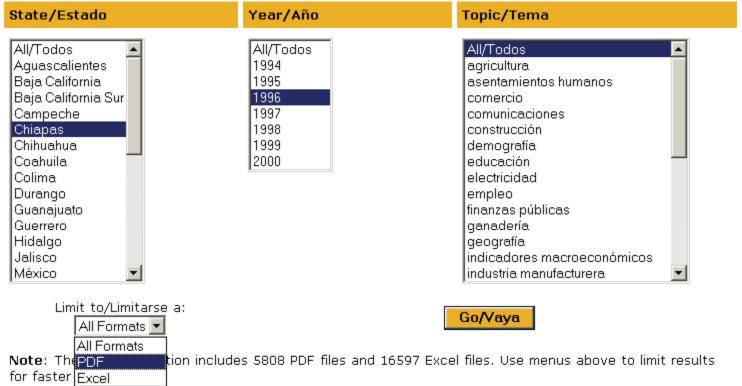
- Select files by year, state, topic, and/or type of file (PDF or Excel)
- Reconstruct the full volume
- Based upon Dublin Core metadata



Help | EGCDL Staff | About EGCDL

Files in this collection were digitized from the series *Anuarios Estatales*, Instituto Nacional de Estadística, Geografía e Informática.

#### Browse Chapters and Tables/Revisar Capítulos y Tablas



Nota: La colección entera incluye 5808 archivos PDF y 16597 archivos Excel. Use los menús para limitar los resultados y expeditar la búsqueda.

ACADEMIC MEDIA & TECHNOLOGY
SOCIAL SCIENCE RESEARCH SERVICES

#### YALE UNIVERSITY LIBRARY

SOCIAL SCIENCE LIBRARIES AND INFORMATION SERVICES

## Economic Growth Center Digital Library

Browse Collection | Help | EGCDL Staff | About EGCDL

Digitized materials are from the series: *Anuarios Estatales*, Instituto Nacional de Estadística, Geografía e Informática

You searched: date:(1996) state:("Chiapas") format:pdf and returned 27 results

More information about PDF files and Excel spreadsheets

Click column headers to sort. Return to default sort.

State/ Estado		<u>Chapter Title/</u> <u>Título Del Capítulo</u>	<u>Chapter Number/</u> Número Del Capítulo	File Type/	File Size/
Chiapas	1996	<u>Indice</u>	01.indice	PDF	55.6 KB
Chiapas	1996	<u>Aspectos Geográficos</u>	01	PDF	2.9 MB
Chiapas	1996	<u>Asentamientos Humanos</u>	02.01	PDF	1.7 MB
Chiapas	1996	Comunicaciones y Transporte	02.02	PDF	1.9 MB
Chiapas	1996	Indice	02.indice	PDF	175.1 KB
Chiapas	1996	<u>Demografía</u>	03.01	PDF	2 MB
Chiapas	1996	<u>Empleo</u>	03.02	PDF	1.6 MB
Chiapas	1996	Salud	03.03	PDF	1.3 MB
Chiapas	1996	<u>Educación</u>	03.04	PDF	1.5 MB
Chiapas	1996	Seguridad y Orden Público	03.05	PDF	1.5 MB
Chiapas	1996	Indice	03.indice	PDF	346.5 KB
Chiapas	1996	<u>Agricultura</u>	04.01.01	PDF	1.7 MB
Chiapas	1996	<u>Ganadería</u>	04.01.02	PDF	1.4 MB
Chiapas	1996	<u>Silvicultura</u>	04.01.03	PDF	866.8 KB
Chiapas	1996	<u>Pesca</u>	04.01.04	PDF	444 KB
Chianas	1996	Minera	N4.N2.N1	PDF	403.1 KB

#### **CAPTACION DE RECURSOS DE LA BANCA** COMERCIAL POR PRINCIPALES INSTRUMENTOS **SEGUN PRINCIPALES LOCALIDADES** Al 30 de septiembre de 1995 (Miles de pesos)

**CUADRO 4.3.3.3** 

	TOTAL		CAPTACION TRADICIONAL				BANCA		
LOCALIDAD		CUENTAS DE CHEQUES	CUENTAS DE AHORRO	DEPOSITOS RETIRA- BLES EN DIAS PREESTA- BLECIDOS	DEPOSI- TOS A PLAZO FIJO	PAGARES CON RENDI- MIENTO LIQUIDABLE AL VENCIMIEN- TO	CAPTA- CION DE RECUR- SOS EN MONEDA EXTRAN- JERA b/	OTROS INSTRU- MEN- TOS	DE INVER- SION a/
ESTADO	3 198 679	1 144 771	20 693	68 285	313 233	1 487 567	15 673	132 750	15 707
TUXTLA GUTIERREZ	1 577 369	571 650	8 385	38 701	194 415	704 359	4 112	41 116	14 631
TAPACHULA	448 960	164 946	3 759	4 254	49 950	189 798	7 045	28 132	1 076
SAN CRISTOBAL DE LAS CASAS	169 200	44 969	2 205	2 773	-	112 481	494	6 278	
HUIXTLA	107 569	35 663	1 327	749	11 965	53 402		4 463	
RESTO DE LOCALIDADES	895 581	327 543	5 017	21 808	56 903	427 527	4 022	52 761	

NOTA: Las cifras de captación de recursos incluyen las operaciones realizadas con el público, atendiendo a la localidad en que se genera el ahorro, y excluyen las efectuadas por agencias y sucursales de bancos mexicanos en el exterior. La captación tradicional y la de banca de inversión se refleren a datos brutos, por tanto, no se elimina la piramidación.

a/ Incluye los saldos de operaciones documentadas como tidelcomisos.

b/ La moneda extranjera está convertida al tipo de cambio del día último del mes, utilizado para solventar obligaciones denominadas en esa moneda, pagaderas









Edit View Favorites Tools Help

ACADEMIC MEDIA & TECHNOLOGY
SOCIAL SCIENCE RESEARCH SERVICES

#### YALE UNIVERSITY LIBRARY

SOCIAL SCIENCE LIBRARIES AND INFORMATION SERVICES

## Economic Growth Center Digital Library

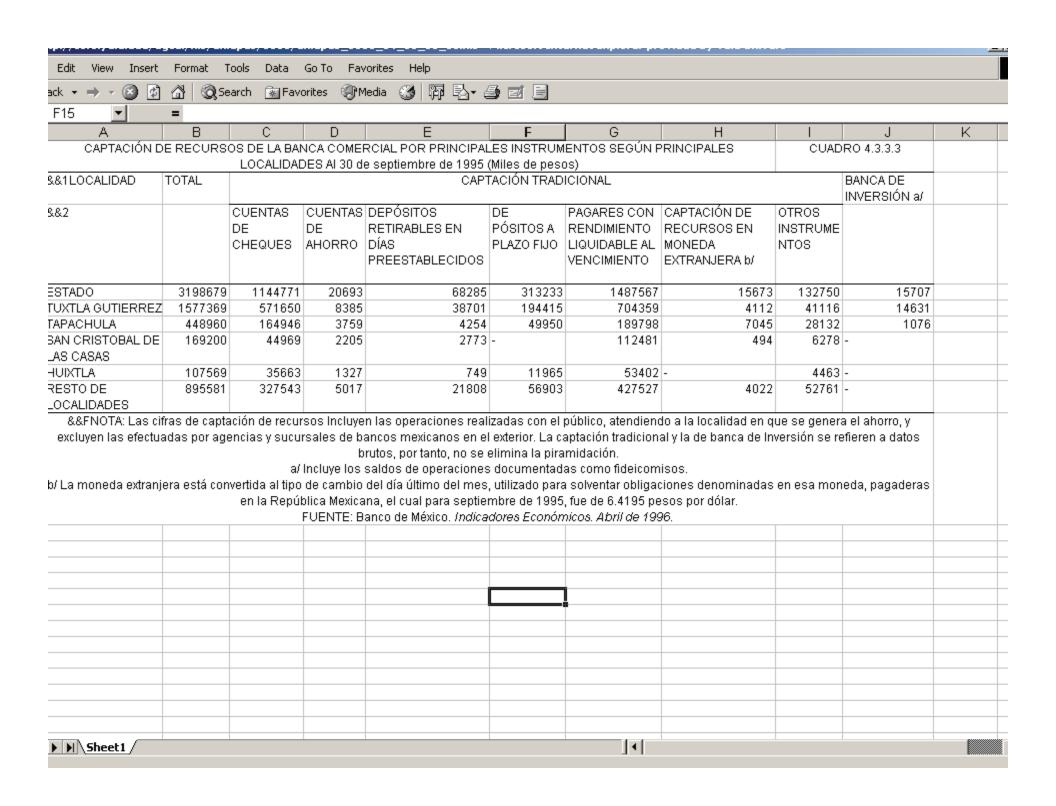
Browse Collection | Help | EGCDL Staff | About EGCDL

Digitized materials are from the series: *Anuarios Estatales*, Instituto Nacional de Estadística, Geografía e Informática

You searched: date:(1996) state:("Chiapas") subject:("servicios financieros") and returned 8 results More information about <u>PDF files</u> and <u>Excel spreadsheets</u>

Click column headers to sort. Return to default sort.

State/ Estado	Year/ Año	Chapter Title/ Título Del Capítulo	<u>Chapter</u> <u>Number/</u> <u>Número Del</u> <u>Capítulo</u>	File Type/ Tipo De Archivo	Table Number (Excel files only)/ Número de la tabla (Excel solamente)	File Size/ Tamaño Del Archivo
Chiapas	1996	<u>Servicios</u> <u>Financieros</u>	04.03.03	Excel	04.03.03.01	19 KB
Chiapas	1996	<u>Servicios</u> <u>Financieros</u>	04.03.03	Excel	04.03.03.02	18 KB
Chiapas	1996	<u>Servicios</u> <u>Financieros</u>	04.03.03	Excel	04.03.03.03	19 KB
Chiapas	1996	<u>Servicios</u> <u>Financieros</u>	04.03.03	Excel	04.03.03.04	18.5 KB
Chiapas	1996	<u>Servicios</u> <u>Financieros</u>	04.03.03	Excel	04.03.03.05-a	19 KB
Chiapas	1996	<u>Servicios</u> <u>Financieros</u>	04.03.03	Excel	04.03.03.05-b	20.5 KB
Chiapas	1996	<u>Servicios</u> <u>Financieros</u>	04.03.03	Excel	04.03.03.06	24 KB
Chiapas	1996	Servicios Financieros	04.03.03	PDF		584.5 KB



## **Next generation interface: NESSTAR**

- Features:
  - Search across tables
  - View tables, select columns and rows
  - Create graphs and charts
  - Download and extract
- Based upon the Data Documentation Initiative (DDI) metadata standard for statistical tables
- Nesstar is in use by data archives in Europe,
   World Bank, Health Canada; in review by ICPSR

## **Nesstar process (1)**

Script creates DDI XML file from Dublin Core record and Nesstar Publisher-ready CSV file from Excel table

EDAD Al 5 de noviembre de 1995	E	F	G H	
CUADRO 3.1.1   POBLACIÓN TOTAL POR SEXO SEGÚN GRUPO QUINQUENAL DE EDAD AI 5 de noviembre de 1995				
TOTAL   HOMBRES   MUJERE				
NTAL         3584786         1790580         1794           ENORES DE UN AÑO         93303         47738         45           A 4 AÑOS         408365         206167         202           A 9 AÑOS         509161         257540         251           A 14 AÑOS         473016         239795         233           A 19 AÑOS         417806         206179         211           A 24 AÑOS         339774         163857         175           A 29 AÑOS         273873         131357         142           A 34 AÑOS         222885         109684         113           A 39 AÑOS         207954         102135         105           A 44 AÑOS         151151         76939         74           A 49 AÑOS         124169         62898         61           A 54 AÑOS         96273         49650         48           A 59 AÑOS         71025         36280         34           A 64 AÑOS         61519         31695         29           A 64 AÑOS         61519         31695         29           A 64 AÑOS         116888         59947         56           D ESPECIFICADO a/         17624         8719         8				
ENORES DE UN AÑO         93303         47738         45           A 4 AÑOS         408365         206167         202           A 9 AÑOS         509161         257540         251           A 14 AÑOS         473016         239795         233           A 19 AÑOS         417806         206179         211           A 24 AÑOS         339774         163857         175           A 29 AÑOS         273873         131357         142           A 34 AÑOS         222885         109684         113           A 39 AÑOS         207954         102135         105           A 49 AÑOS         151151         76939         74           A 49 AÑOS         124169         62898         61           A 54 AÑOS         96273         49650         48           A 59 AÑOS         71025         36280         34           A 64 AÑOS         61519         31695         29           A 64 AÑOS         61519         31695         29           A 64 AÑOS         116888         59947         56           A 65 AÑOS Y MAS         116888         59947         56           A 59 ESPECIFICADO a/         8719         8719 <t< td=""><td>8</td><td></td><td></td><td></td></t<>	8			
A 4 AÑOS				
A 9 AÑOS       509161       257540       251         A 14 AÑOS       473016       239795       233         A 19 AÑOS       417806       206179       211         A 24 AÑOS       339774       163857       175         A 29 AÑOS       273873       131357       142         A 34 AÑOS       222885       109684       113         A 44 AÑOS       207954       102135       105         A 49 AÑOS       151151       76939       74         A 49 AÑOS       124169       62898       61         A 54 AÑOS       96273       49650       46         A 64 AÑOS       71025       36280       34         A 64 AÑOS       61519       31695       29         E AÑOS Y MAS       116888       59947       56         D ESPECIFICADO a/       17624       8719       8         &Fa/l Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".       50	565			
0 A 14 AÑOS       473016       239795       233         6 A 19 AÑOS       417806       206179       211         0 A 24 AÑOS       339774       163857       175         6 A 29 AÑOS       273873       131357       142         0 A 34 AÑOS       222885       109684       113         6 A 39 AÑOS       207954       102135       105         0 A 44 AÑOS       151151       76939       74         6 A 49 AÑOS       124169       62898       61         0 A 54 AÑOS       96273       49650       46         0 A 54 AÑOS       96273       49650       46         0 A 64 AÑOS       71025       36280       34         0 A 64 AÑOS       61519       31695       29         5 AÑOS Y MAS       116888       59947       56         0 ESPECIFICADO a/       17624       8719       8         8Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".       5         FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien	198			
A 19 AÑOS       417806       206179       211         A 24 AÑOS       339774       163857       175         A 29 AÑOS       273873       131357       142         A 34 AÑOS       222885       109684       113         A 44 AÑOS       207954       102135       105         A 44 AÑOS       151151       76939       74         A 49 AÑOS       124169       62898       61         A 54 AÑOS       96273       49650       48         A 59 AÑOS       71025       36280       34         A 64 AÑOS       61519       31695       29         AÑOS Y MAS       116888       59947       56         D ESPECIFICADO a/       17624       8719       8         &Fa/Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".       FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien				
A 24 AÑOS       339774       163857       175         A 29 AÑOS       273873       131357       142         A 34 AÑOS       222885       109684       113         A 39 AÑOS       207954       102135       105         A 44 AÑOS       151151       76939       74         A 49 AÑOS       124169       62898       61         A 54 AÑOS       96273       49650       46         A 59 AÑOS       71025       36280       34         A 64 AÑOS       61519       31695       29         A MOS Y MAS       116888       59947       56         D ESPECIFICADO a/       17624       8719       8         &Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".       FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien				
A 29 AÑOS       273873       131357       142         A 34 AÑOS       222885       109684       113         A 39 AÑOS       207954       102135       105         A 44 AÑOS       151151       76939       74         A 49 AÑOS       124169       62898       61         A 54 AÑOS       96273       49650       46         A 59 AÑOS       71025       36280       34         A 64 AÑOS       61519       31695       29         AÑOS Y MAS       116888       59947       56         DESPECIFICADO al       17624       8719       8         &Fal Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".       50				
A 34 AÑOS       222885       109684       113         A 39 AÑOS       207954       102135       105         A 44 AÑOS       151151       76939       74         A 49 AÑOS       124169       62898       61         A 54 AÑOS       96273       49650       46         A 64 AÑOS       71025       36280       34         A 64 AÑOS       61519       31695       29         AÑOS Y MAS       116888       59947       56         DESPECIFICADO a/       17624       8719       8         &Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".         FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien				
6 A 39 AÑOS       207954       102135       105         1 A 44 AÑOS       151151       76939       74         6 A 49 AÑOS       124169       62898       61         1 A 54 AÑOS       96273       49650       48         6 A 59 AÑOS       71025       36280       34         1 A 64 AÑOS       61519       31695       29         6 AÑOS Y MAS       116888       59947       56         D ESPECIFICADO a/       17624       8719       8         &Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".         FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien				
0 A 44 AÑOS       151151       76939       74         5 A 49 AÑOS       124169       62898       61         0 A 54 AÑOS       96273       49650       46         5 A 59 AÑOS       71025       36280       34         0 A 64 AÑOS       61519       31695       29         5 AÑOS Y MAS       116888       59947       56         0 ESPECIFICADO a/       17624       8719       8         &Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".         FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien.				
6 A 49 AÑOS12416962898611 A 54 AÑOS9627349650486 A 59 AÑOS7102536280341 A 64 AÑOS6151931695296 AÑOS Y MAS11688859947565 ESPECIFICADO a/1762487198&Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien				
A 54 AÑOS 96273 49650 4665 5 A 59 AÑOS 71025 36280 349 AÑOS 61519 31695 29 AÑOS Y MAS 116888 59947 566 5 AÑOS 116888 59947 566 5 AÑOS Y MAS 116888 566 5 AÑOS Y MAS 11688 566 5 AÑOS Y MAS 116888 566 5 AÑOS Y MAS 11688 566 5 AÑOS Y MAS 116888 566 5 AÑOS Y MAS 11688	212			
5 A 59 AÑOS 71025 36280 34 0 A 64 AÑOS 61519 31695 29 5 AÑOS Y MAS 116888 59947 56 0 ESPECIFICADO a/ 17624 8719 8 8Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes". FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien	271			
DA 64 AÑOS 61519 31695 29  5 AÑOS Y MAS 116888 59947 56  O ESPECIFICADO a/ 17624 8719 8  &Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes".  FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien	623			
i AÑOS Y MAS 116888 59947 56 D ESPECIFICADO a/ 17624 8719 8 &Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes". FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien	745			
O ESPECIFICADO a/ 17624 8719 8 &Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes". FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien	824			
&Fa/ Incluye una estimación de población por un total de 13 935 personas, correspondientes a 2 994 "vivien sin Información de ocupantes". FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien	941			
sin Información de ocupantes". FUENTE: INEGI. Chiapas, Resultados Definitivos; Tabulados Básicos; Tomo I. Conteo de Población y Vivien	905			
	da e			
M Sheet1				

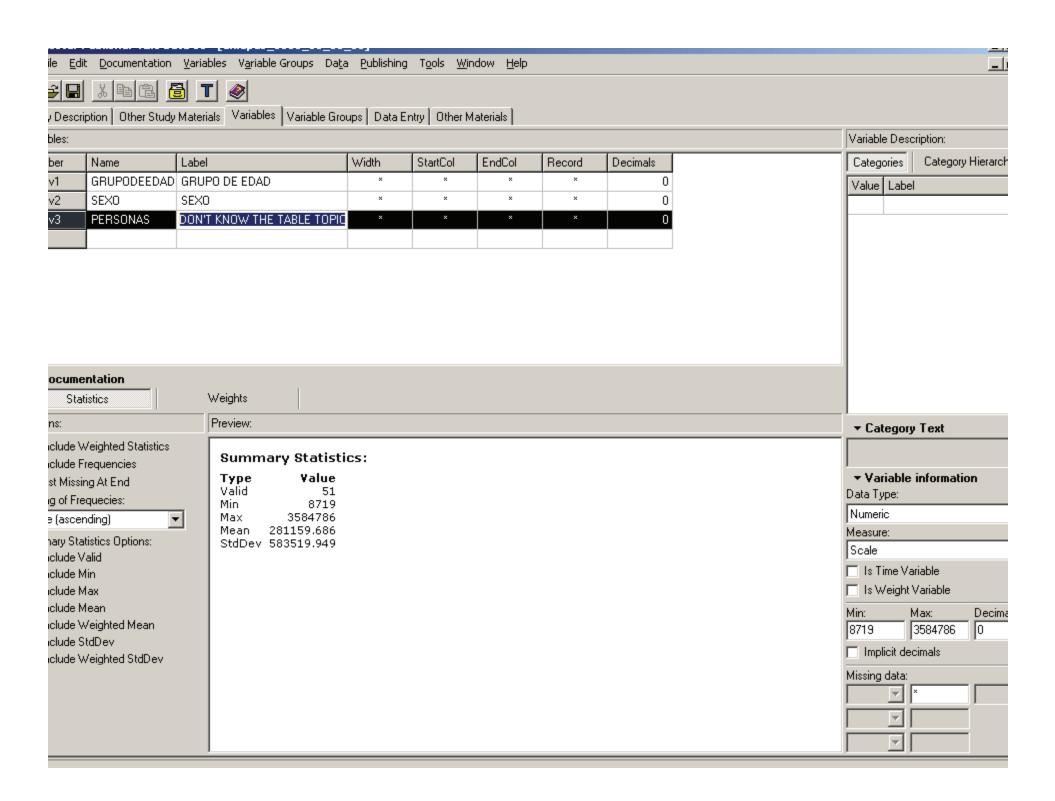
```
Edit Search View Tools Options Language Buffers Help
🚅 🛛 X | 🚭 | ¼ 🖺 🖺 X | ∽ ∼ | Q 🐗
apas_1998_03_01_01.xml
<?xml version='1.0' encoding='utf-8'?>
<metadata xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:title>Estado y Movimiento de la Población</dc:title>
<dc:subject>demografía</dc:subject>
<dc:date>00-00-1998</dc:date>
<dc:format>application/vnd.ms-excel</dc:format>
<dc:format>18 KB</dc:format>
<dc:identifier>http://ssrs.yale.edu/egcdl/xls/Chiapas/1998/Chiapas_1998_03_01_01.xls</dc:identifier>
<dc:source>chia199803.01.01</dc:source>
<dc:source>Anuario Estadístico del estado de Chiapas, 1998</dc:source>
<dc:coverage>Chiapas</dc:coverage>
<dc:coverage>mx</dc:coverage>
</metadata>
```

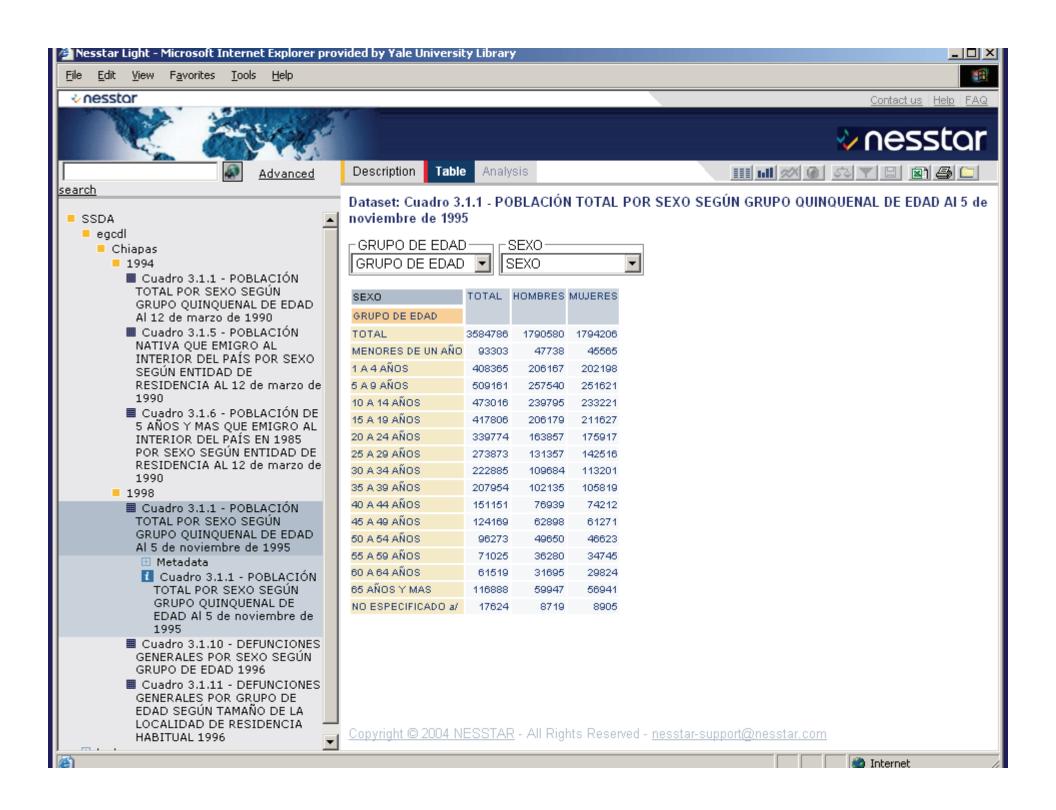
## **Output from the script**

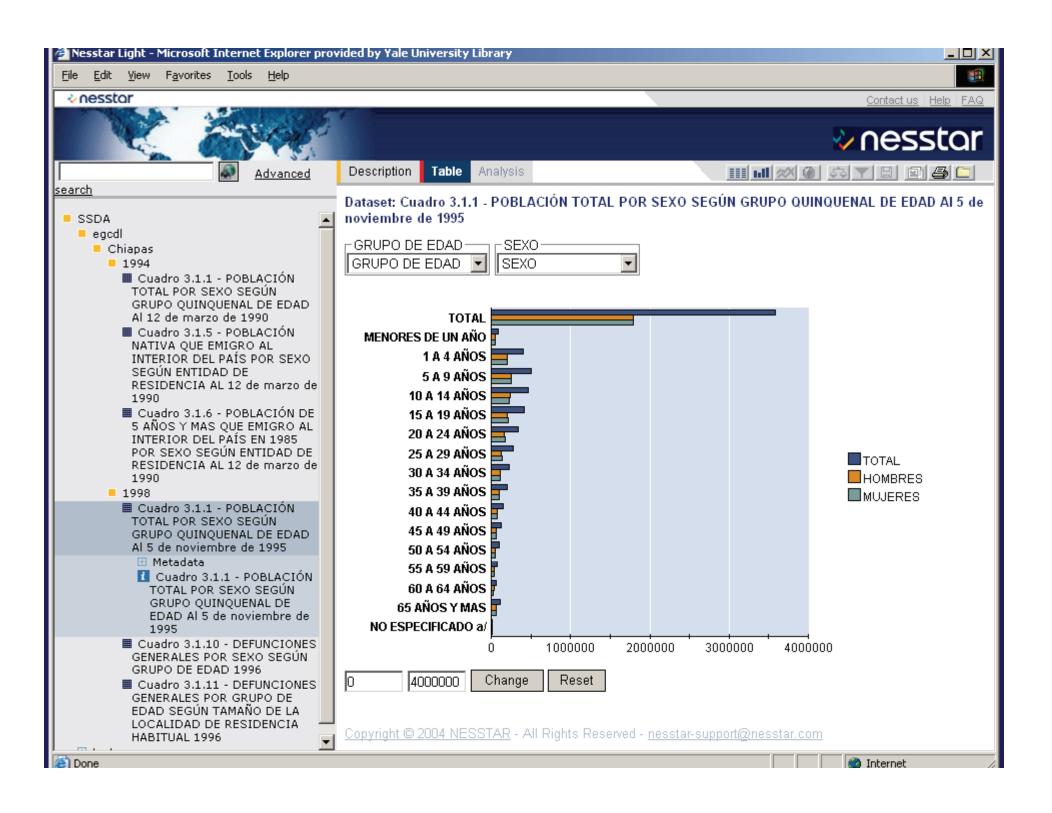
- **#CSV** file formatted for Nesstar Publisher
- **#DDI** record to import into Publisher

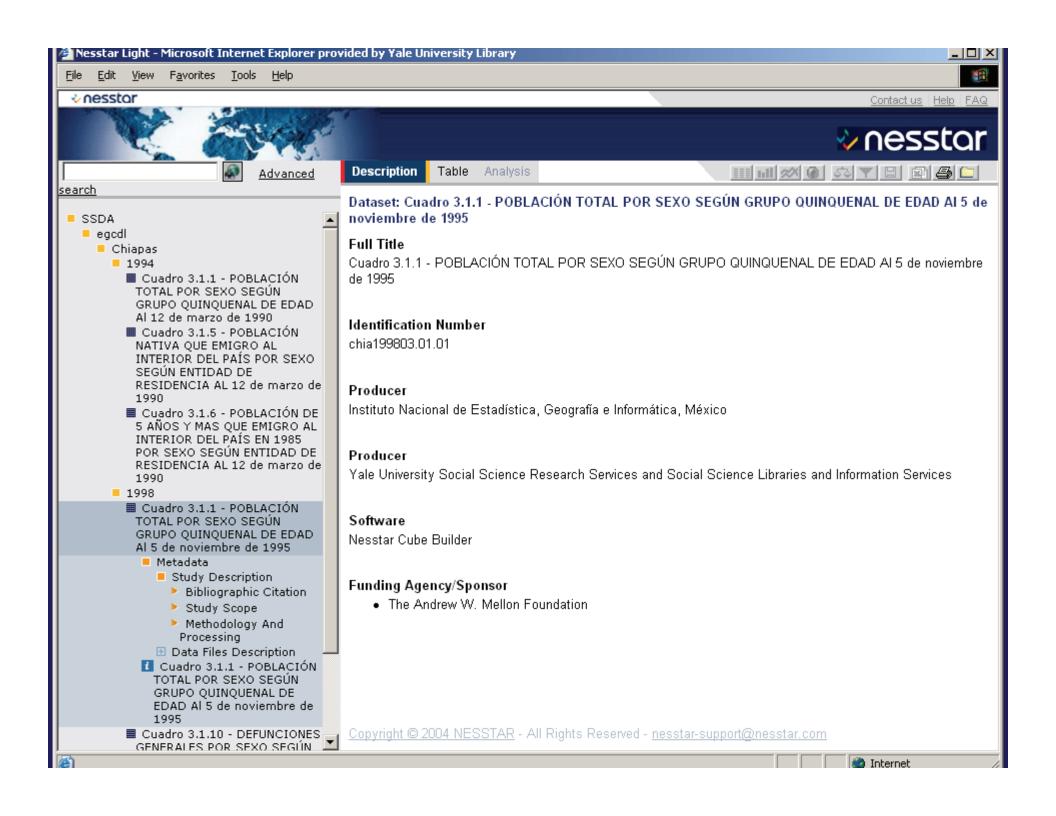
## **Nesstar process (2)**

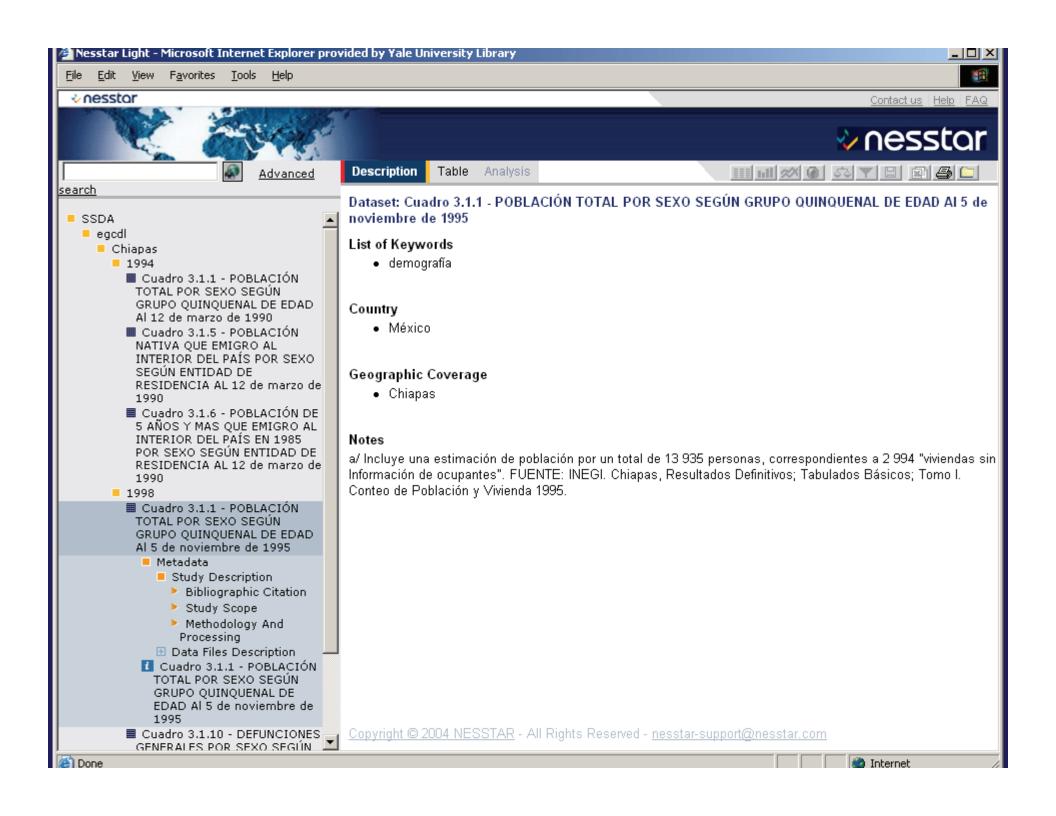
- ## Staff will publish each pair of DDI XML file and Excel table using Nesstar 'cube builder' software
- **#Some editing needed at this stage:** 
  - Add 'measure' (what the table is measuring—persons, events, pesos, etc)
  - Add column header name
- #Publish metadata and table data in Nesstar server

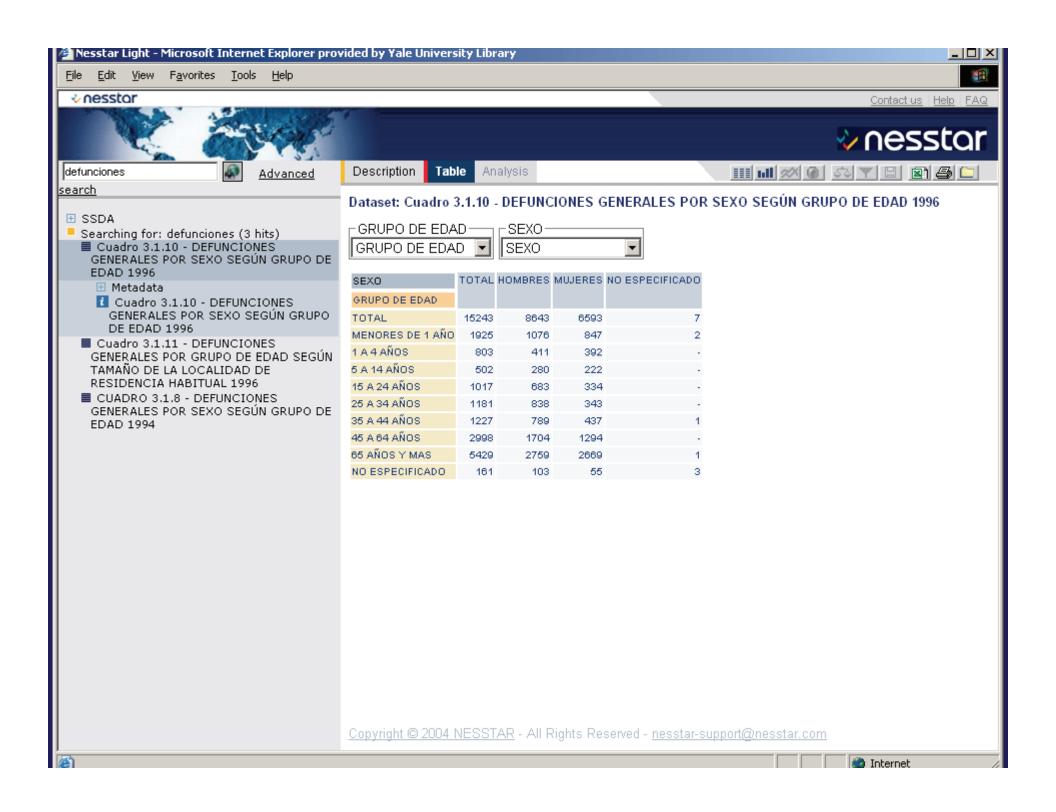












### **Next 6 months**

- **\*\*** Nesstar implementation and evaluation
- **#**Analysis of research questions
- **\*\***Local usability review of PDF, Excel, and Nesstar interfaces
- # External review committee visit in July
- **\*\*Numerous presentations at professional** meetings
- **#**Reporting to Mellon