

GCBS Exchanges with David Damas

Woodrow W. Denham and David Damas, 2002

The following is the first set of excerpts from letters I exchanged with David Damas concerning his Central Arctic datasets in the GCBS Database. I appreciate his sharing his important concerns about my manipulation of the Inuit data.

Contents

2002-04-06 DENHAM TO DAMAS.....	2
2002-04-22 DAMAS TO DENHAM.....	4
2002-04-27 DENHAM TO DAMAS.....	6
2002-05-09 DENHAM TO DAMAS.....	7
2002-06-21 DAMAS TO DENHAM.....	16
2002-08-30 DENHAM TO DAMAS.....	19
2002-09-28 DENHAM TO DAMAS.....	20
2002-09-29 DENHAM TO DAMAS.....	22

2002-04-06 Denham to Damas

To: David Damas, Burlington, Ontario, CA
From: Woody Denham, Sharjah, UAE
Date: 6 April 2002
Subj: Central Arctic populations

Dear David,

. . . . About a year ago, I dug out all of the many projects that I put on hold in 1982. Using technology that did not exist twenty years ago, I have done a fairly effective job of resurrecting and upgrading them to the standards of 2002. I have not finished yet and in fact I plan to do a great deal more in coming months and years, but I am far enough along with the work that I want to share it with you.

I'm sure you remember all of the work you and I did in 1978-80 to enhance the genealogies and census data from the 5th Thule Expedition and your later work with the Iglulik. The fact that I did not published those data years ago is one of the great casualties of my dropping out of academic anthropology. However, I am writing to you now because I have recently made all of my band compositions data available on the Internet, including everything we assembled from the Central Arctic in 1978-80.

I am working closely albeit entirely electronically with Dr. Douglas White at the University of California, Irvine, who is interested not only in the band compositions data but also in my Alyawarra data from Central Australia. I have recently published the first installment of the Group Compositions in Band Societies (GCBS) Data Base on Doug's server at UCI.

GCBS contains 41 datasets from band societies around the world, including 13 Inuit datasets stretching from Angmagsalik to Point Hope. Most importantly it contains all of your enhanced and upgraded Central Arctic data including: Iglulik 1921-22, Iglulik 1949, Iglulik 1961, Iglulik 1960-61 Winter Seasonal Cycle, Netsilik 1922-23 and Copper 1922-23. The attached statistical summary provides a comprehensive overview of the GCBS Data Base, including your data.

I say the first installment of GCBS is on the web now for there is more to come. All of the numerical data are up now, as are all of the genealogical diagrams and other supporting files for the 13 Inuit datasets. But I am not pleased with the appearance of the genealogical diagrams, so I am converting a completely new set of diagrams for all 41 datasets to a suitable electronic format. They should be on the web by the end of April. In addition to the numerical and graphical data, GCBS also contains a user manual and other supporting files to enhance the usability of the datasets.

The URL for the GCBS Data Base is:

If you can provide an email address that I can use to send the URL to you in an electronic format, it will be a lot easier to handle than it is in this paper copy.

I would be delighted to have you look at all of the GCBS materials, focusing especially on the Central Arctic datasets, and give me some feedback. The large numerical data files containing the genealogies and censuses do not make easy reading, but all is not lost. The user guide briefly describes the database as a whole and the general features of all datasets, the “Key” files contained in the datasets describe the details that are unique to the sets, and the genealogical diagrams for the Inuit sets – aesthetically imperfect though they may be – are as accurate as you and I could make them.

I am mortified by the fact that it took me over 20 years to make your files available to the world at large, but derive some tiny bit of consolation from the old adage “better late than never”.

I shall be doubly mortified if you examine the data base and discover serious or fatal flaws in it, but if you find anything I should correct, expand or clarify, please let me know. Both the GCBS project and my similarly resurrected Alyawarra project are “works in progress” and will remain in that status permanently. One of the joys of publishing on the web is the freedom to correct errors whenever I find them.

Woody Denham

* * * * *

2002-04-22 Damas to Denham

From: David Damas, Burlington, Ontario, CA
To: Woody Denham, Sharjah, UAE
Date: 22 April 2002
Subj: Central Arctic populations

. Turning to the material which you sent, I much prefer to work with diagrams than with webware. I haven't adapted to the computer age. The closest that I have come to technological innovation is use of an electronic typewriter and the FAX system.

I am rather puzzled at the categories which you have used in your summaries. The "number of living people" column is clear enough, but note my comments below on "squared" individuals. Does "number of records" refer to either or both of the following (1) deceased alone; or (2) those and living but absentee members? These have been included in my diagrams in order to show kinship based membership in addition to those linked by existing ties. I find the greatest difficulty with the next column where you seem to equate camps- villages (this is O.K.) with "households." I don't think we can do this. In the place, camps or villages are local groupings while "households" can be constituent units of such or, more rarely the whole local group.

Then there is the problem of defining households. Generally we have been using this term to indicate people domiciled together, either (1) those under one roof; or (2) those sharing a close knit compound of dwellings. Such units can be either (1) nuclear families or (2) some form of composite or extended family. Only rarely unrelated persons. These units cannot ordinarily be abstracted from diagrams of group composition unless the anthropologist has used some dividing symbol or feature to designate them (see discussion below). For this column, I would suggest the convention employed by June Helms "marital pairs" as used in her statistical formulation of band composition.

I have not had time to go through all of my Central Eskimo charts, but I have reexamined those for Iglulik in 1921-1949 and 1960-61, and am puzzled by some of your summarizing results. Turning first to 1921-22, your number of "living people" is about right if you exclude living but absent (the circles and triangles within squares). For total number of records, if I read you correctly to include the above and deceased I get 164 rather than your 178. (less if you exclude living but absent). In the next column you are clearly counting number of marital pairs which I think should be consistently used here. In the 1949 census, however you are counting number of local groups, at 11 while marital pairings would be 66. Also, for 1949 I get 316 to your 334 if we are using the same criteria for "total records". In the 1960-61 census your number of living and present people seems right. I haven't taken the time to count "records" because I wanted to get this off to you quickly. However, your number in the next column is really strange. I think you are as in 1949 counting local groups but I get 16 rather than your 20. Total marital would be 103. These correspond to number of nuclear families which

raises again the question of households, since I found that I had to group 78 of these units into extended family households (including compounds). This situation is in marked contrast to the Copper Eskimo picture where the isolated nuclear family comprised most households or domiciled units. The group composition charts of the Copper Eskimo are virtually indistinguishable from those of Iglulik, so that this contrast in family organization is not shown on them.

I hate to be so qualified in my comments but since you are webbing both the charts and your summaries of them it is important that there is agreement. Unless of course where I am finding disagreement there is really some misunderstanding on my part.

David Damas

* * * * *

2002-04-27 Denham to Damas

To: David Damas, Burlington, Ontario, CA
From: Woody Denham, Sharjah, UAE
Date: 27 April 2002
Subj: Central Arctic populations

Dear David,

I have read your letter carefully and believe that most if not all of the problems you describe are based on differences in definitions, formatting and counting. I simply don't have time to respond carefully right now because of the midterm exams, but will do so as soon as possible.

In the meantime, I notice that your comments deal primarily with the Data Summary table and the Genealogical Diagrams, but I do not see any reference to your having examined the Database Documentation file or the Numerical Data files.

The data that appears in the Summary comes directly from the Numerical Data files which contain a great deal more information than appears in the Genealogical Diagrams. I understand that a picture is worth a thousand words and for that reason I included the genealogical diagrams with the datasets, but the "real" data is in the numerical files, not in the diagrams. When I have time to write more in a few days, I'll deal almost exclusively with the numerical data, so if you did not examine it or make a copy of it, you might want to do so. Also, I suggest you take a look at the data summary that appears in the Key file for each dataset.

Regards, Woody

2002-05-09 Denham to Damas

To: David Damas, Burlington, Ontario, CA
From: Woody Denham, Sharjah, UAE
Date: 9 May 2002
Subj: Central Arctic populations

Dear David,

> *Turning to the material which you sent, I much prefer to working with diagrams
> than with webware. I haven't adapted to the computer age. The closest that I
> have come to technological innovation is use of an electronic typewriter and
> the FAX system (note my contact at a local Mail Boxes Etc. station below).*

I understand your reluctance to make the transition. I too am a long way behind the so-called "leading edge". But I enjoy working with good technology that works well. I would be paralyzed without my computer and scanner.

My opportunity to put the database on the web came along before all of the files were ready, so I put them up ASAP with an expectation that I would have to make changes later. The main reason for my delay in responding to you has been the intensive effort I have made in recent days to complete the current round of revisions to the GCBS Database. These revisions are not on the web yet but will be in a few days I hope. They include but are not limited to the following:

- **Genealogical diagrams.** Until yesterday the Inuit datasets contained poor quality copies of the diagrams, but now they contain new, clean copies of them. Until yesterday, the other datasets did not contain any genealogical diagrams at all, but now all of them contain complete sets of diagrams. Also I sharply upgraded the quality of the maps in the Inuit datasets.
- **Data Summary.** The summary table that lies at (or near) the heart of your comments was not fully documented. Now I have prepared a key that appears at the top of the summary page and briefly explains each column header. I include a copy of the key with this letter.

Overview of the GCBS Data Base.

I suspect that some of the problems you experienced in your first encounter with the Database were related to my rather cryptic descriptions in a few places – perhaps so cryptic that they were invisible. Let me try again. I appreciate having this opportunity to expand, clarify and revise, which of course is one of the great joys of publishing on the web.

When you access a specific dataset from the Dataset Table of Contents, for example Iglulik 1949, you immediately display the Key file for that dataset. The Key contains four kinds of items:

- a clickable table of contents for all of the remaining files in the dataset
- a text description of the dataset, including a brief or expanded description of the genealogical diagrams, and anything about the organization or coding of the dataset that is unique to that set (remember that generic descriptions that apply to all datasets appear in the Manual).
- an explanatory key to unique features of the numerical codes that are used in this specific numerical data file.
- a very brief set of summary statistics concerning various types of residential groups represented in the dataset (Inuit datasets only).

The Key file, like GCBS as a whole, is organized so you can study the items in the Key file the first time you display it, then skip the text and go directly to the data on subsequent visits to the site.

The clickable table of contents that appears at the top of each Key file is organized as follows:

- Numerical data
- Genealogical diagrams (at least 1, as many as 17)
- Map (Inuit only)

I discuss these options in reverse order, again operating on the assumption that a person would look at the map once and ignore it on subsequent visits to the site, look at the genealogical diagrams now and then as needed, but focus primarily on the numerical data that appears at the top of the list.

The *map* has no scientific value. It illustrates roughly where the society is located, and in general where the groups depicted in the genealogical diagrams are located. It is included for orientation, not for analysis. The new version is better than the old, but the Netsilik and Copper maps are not finished yet.

The *genealogical diagrams* are included primarily as illustrations of many of the relationships contained in the numerical data files, and serve the important function of increasing redundancy and facilitating error correction within the dataset. However, the diagrams, unlike the map, contain a good bit of analyzable *secondary* data concerning topics such as adoption that are not coded in the numerical data files, so the diagrams are not *just* illustrations. The kinds of secondary data they contain are explained in the Manual and in the individual Key files, and vary from set to set.

The diagrams are *not* primary data in GCBS for they are much too limited in scope and flexibility to do justice to your extraordinarily rich Inuit data, to my Alyawarra data, and in fact to many of the datasets included here. Certainly there should be no disagreements between the diagrams and the numerical data, but the numerical data files contain

vastly more information than anybody could ever collapse into a single diagram, so the diagrams play supporting roles here and are not the stars of the show.

Finally we arrive at the **numerical data file** for each dataset. This is where the analyzable data resides. The **Manual** describes the numerical data file layout in considerable detail, including a general description of the *variables* that appear in the columns and the *values* that appear in the cells. The **Key** file for each dataset contains a specific description of the variables and values unique to that dataset. Each record in the numerical data file pertains to one person, and the number of records in the files ranges from a low of 19 (AU04 Ngatatjara) to a high of 704 (AF02 Mbuti Forest).

Responses to your concerns

In responding to your comments, it is unfortunate that I don't have your data in its published form with me here in the Middle East. I make the following comments 20+ years after I did the coding and without the assistance of any supporting documentation. If I get something wrong, please help me correct it.

Model diagrams. Generally speaking, as I prepared diagrams for the 41 datasets in GCBS I followed existing diagrams when they looked more or less like the "standard" that emerged as I did the project, but often that policy failed. I simply don't remember how closely I followed the layout of your diagrams when I constructed the ones in GCBS, but I assume the two sets contain a number of (perhaps significant) differences. But if the logic is the same even though the appearances are different, we probably are OK.

> *I have not had time to go through all of my Central Eskimo charts, but I have reexamined those for Iglulik in 1921, 1949 and 1960-61, and am puzzled by some of your summarizing results. Turning first to 1921-22, your number of "living people" is about right if you exclude living but absent (the circles and triangles within squares). For total number of records, if I read you correctly to include the above and deceased I get 164 rather than your 178. (less if you exclude living but absent).*

Missing people. Your published **texts** contained references to some people who did not appear in your **diagrams**. I know your diagrams included several people who were absent, but I seem to recall that I found a few more references in the text to people that were not in your diagrams.

Generally speaking I represented these "missing people" in the diagrams with a filled triangle or circle *without* the X-mark used to indicate deceased. So an open O is a living female who was part of the censused population, a filled or closed O is a living female who lived outside of your research population, and a filled O with a superimposed X was a deceased female. In the genealogical diagram for the dataset called NU08 Iglulik 1960-61 Seasonal Cycle, notice that woman #85 was such a person – a widow living

elsewhere who was a link between her sons #28, 30 and 31. In the numerical data files, people like this are shown with Life Status = 1 Alive, but there is no census data for them.

I have not checked to see how many people in each dataset were in this category and without my notes I cannot distinguish now between the ones you included and the additional ones that I found. I am sure the numbers were small even though the connections established by those people may have been important. In any event, I am pretty sure I included several such people in my tables and diagrams who did not appear in the original diagrams.

Square people. I used squares to represent people of unknown or ambiguous sex. Unknown simply means unstated in the source; ambiguous means a contradiction in the source such as a diagram that shows a female accompanied by a statement in the text that refers to a male. In these cases I included these people (usually young children) on the diagrams and in the numerical data file, I assigned ID# to them and showed them as living, etc. But sex is marked as unknown. So if you count the total number of living people in the population, these people are counted, but for example if you compute separate totals for *males* and *females*, they are totaled in the *unknown* category.

In some datasets squares are used without ID# in the diagrams. In such cases, a square represents an unknown number of children of unknown sex. Since the source simply indicated that a man and a woman had unspecified children, I indicated that in the genealogical diagrams with an unmarked square, but such individuals do not appear in any way in the numerical data files. Since we don't know how many or which sex (or how old), we really can't say much about the children as individuals in a census. This may not apply to the Inuit sets, but it is a serious problem in some other datasets.

Deceased people. In some datasets (I don't remember the specifics of the Inuit sets), it was possible for me to infer the existence of deceased people who did not appear in the genealogies and diagrams. For example, if two people were ***unambiguously*** identified in the text as full brothers or as half brothers with the same mother, etc., but the parents of those two men were not identified further in the text or diagrams, I treated that situation as an unintended omission by the writer and inserted the missing links into the numerical data files and the diagrams. The addition of such ***unambiguously*** identified links has no bearing on the census data for these additions represent deceased people, but they slightly increase the genealogical richness of the datasets. I emphasize that I inserted these missing links ***only*** when the source was ***absolutely unambiguous***. If there was even a shadow of a doubt, I didn't add anything. The result of my judicious addition of these missing links is that a count of the deceased ancestors in the numerical data files may yield a slightly higher number of deceased people than you would find by looking at the diagrams in the sources.

Total Number of Records. The Total Number of Data Records in each numerical data file is the sum of the following:

- living people who lived within the research population.

- “square people” (alive by definition) when we at least know how many square people to include.
- “missing people” who were stated to be alive but lived elsewhere.
- deceased people (including the “missing links” that I added).

So the difference between Total Number of Data Records and Total Living People (including “square” and “missing” people”) is the Total Number of Deceased People represented by records in the numerical data file. That number can be computed by subtracting Total Living People from Total Records, so I omitted it from the Statistical Summary since the page is not wide enough to hold everything. Perhaps my new documentation for the Summary will help here.

Number of Households, Camps and Villages BY Censuses. This column in the statistical summary seems to present the greatest problems, so I deal with it here in some detail.

> I find the greatest difficulty with the next column where you seem to equate
 > camps-villages (this is O.K.) with “households.” I don’t think we can do this. In
 > the place, camps or villages are local groupings while “households” can be
 > constituent units of such or, more rarely the whole local group.

> However, your number in the next column is really strange. I think you are as in
 > 1949 counting local groups but I get 16 rather than your 20. Total marital would be
 > 103. These correspond to number of nuclear families which raises again the
 > question of households, since I found that I had to group 78 of these units into
 > extended family households (including compounds).

I begin by quoting a paragraph that appears at the beginning of the new version of the Statistical Summary:

Number of Households, Camps and Villages BY Censuses. This figure is a ballpark estimate of the number of censused residential groups represented in each dataset. In the minimal case, if a dataset referred to only one camp at one time and contained no information on household compositions, the number in this column would be “1”. On the other hand, if the set referred to two camps containing a total of nine households, and the report contained two censuses of those groups, the number in this column would be “22” (2 camps + 9 households = 11 residential groups; 2 censuses of 11 residential groups = 22 censused residential groups). Computing the precise number can be challenging, especially in highly fluid situations in which complex changes occur through time in the number of censused villages, camps, households and other residential groups.

To expand upon that statement, I have excerpted some material from the Key files from the various Iglulik datasets.

In the Key to NU05 Iglulik 1921-22, we see the following:

11. ResGrp1 (1-5) Villages, Winter 1921-22
12. ResGrp2 (1-48) Households, Winter 1921-22

Variable (column) 11 is called ResGrp1, it has 5 values (1-5), the 5 units in question are villages, and the census data for those villages pertains to the Winter of 1921-22.

Variable 12 is called ResGrp2, it has 48 values (1-48), the 48 units in question are households, and the census data for those households pertains to the Winter of 1921-22.

I extracted these villages and households from the NU05 Iglulik 1921-22 sources (textual as well as graphical) and coded them in the numerical data file in such a way that the original details are fully preserved. If you want to analyze by village, then focus specifically on data in the ResGrp1 column. If you want to analyze by household, focus specifically on ResGrp2. If you want to analyze by both village and household, then focus on both ResGrp1 and ResGrp2. In other words, all of the data is there for villages **and** households.

When I “do the numbers” in a very preliminary fashion, I find that NU05 Iglulik 1921-22 contains a single census of 5 villages and 48 households, or a total of 53 residential groups that are directly analyzable within the numerical data file. That is the number that appears in the Statistical Summary. It is minimally sophisticated, but serves as an informative albeit imprecise index to the scope and complexity of the dataset. For example, if you look at that column and compare the numbers for Iglulik, Netsilik and Copper with most other data sets, you instantly see that they together constitute a sizable percentage of the entire data base, not just in terms of numbers of people, but also in terms of the richness of the data.

In the Key to NU08 Iglulik 1960-61 Seasonal Cycle, we see the following much more complex situation:

10. ResGrp1 (1-3) Villages, 1 Aug 1960
11. ResGrp2 (1-3) Villages, 12 Aug 1960
12. ResGrp3 (1) Villages, 15 Sep 1960
13. ResGrp4 (1) Villages, 15 Dec 1960
14. ResGrp5 (1-3) Villages, 1 May 1961
15. ResGrp6 (1-3) Villages, 8 May 1961
16. ResGrp7 (1) Villages, 17 May 1961
17. ResGrp8 (1-2) Villages, 1 July 1961
18. ResGrp9 (1) Villages, 31 July 1961
19. ResGrp10 (1-19) Households, 15 Dec 1961

In Variables 10-18 you provided a series of discrete censuses for 1, 2 or 3 villages that were merging and separating and moving about the landscape for an entire year. Since

each one of those censuses is a real “thing” in its own right, this dataset contains 18 village censuses. In addition it contains 19 household censuses in Var19 ResGrp10. Thus the quick and dirty comparative figure for the NU08 Iglulik 1960-61 Seasonal Cycle dataset is 18 villages + 19 households = 37 censused residential groups. That number appears in the Statistical Summary.

Because of all of the changes that occurred within this group of people during the 1960-61 seasonal cycle, the only sensible way to diagram genealogical relations among the people was to generate a single diagram that says nothing at all about residential groupings of any kind.

In the specific case of Iglulik 1961, we see the following:

11. ResGrp1 (1-16) Villages, January-May 1961
12. ResGrp2 (1-4) Neighborhoods in Villages 15 and 16, Winter 1961

Since I don't have the raw data with me, I don't know exactly what ethnographic reality corresponds to the four Neighborhoods in Villages 15-16 as coded in Var12. But the 16 villages coded in ResGrp1 correspond exactly to your count of 16, and the 4 neighborhoods coded in ResGrp2 are in addition to the village count of 16, so we have a total of 20 analyzable residential groups.

> *Then there is the problem of defining households. Generally we have been using this term to indicate people domiciled together, either (1) those under one roof; or (2) those sharing a close knit compound of dwellings. Such units can be either (1) nuclear families or (2) some form of composite or extended family. Only rarely unrelated persons. These units cannot ordinarily be abstracted from diagrams of group composition unless the anthropologist has used some dividing symbol or feature to designate them (see discussion below).*

I extracted each of the group compositions (household, village, etc.) directly from the data in the source (diagrams, tables, texts, etc.) and coded it numerically. If the source did not enable me to make such distinctions, then the distinctions do not appear in the numerical data files. And indeed several of the files contain only a single “level” of data, typically a single census for a single camp or village with no data for households or any other kinds of groups.

> *For this column, I would suggest the convention employed by June Helms “marital pairs” as used in her statistical formulation of band composition.*

Years ago I seriously considered adopting her scheme, but decided not to do it. Basically this is the so-called “splitter-lumper” problem. I almost always come down on the splitter side. If I code things at the simplest / smallest / lowest level of abstraction, it is then easy enough for others to lump them as they see fit; whereas if I lump them at the beginning, it can be difficult or impossible for others to split them for other purposes. Since the ultimate “atom” of the GCBS Database is the individual person, and since

those people can be combined in all sorts of ways that I could not anticipate years ago, I opted for simplicity rather than complexity. Perhaps that makes me a Neo (or Paleo-Neo) Empiricist.

Of course marital pairs are embedded in the numerical data. In the following excerpt from the NU07 Iglulik 1961 numerical data file, we see in the first data record that ID#11 (Var1) is the son of ID#552 (Var6) and ID#553 (Var7), and that he is married to ID#278 (Var8). Likewise, ID#12 is married to ID#282, 280 and 687 (deceased). And so on. If anyone wants to work with marital pairs, all of them are right there.

```
File.Rcrd....1.2.3..4.5...6...7...8...9..10.11.12..13
NU07 0011    11 1 1  0 2 552 553 278    0    0  1  0  1.03
NU07 0012    12 1 1  0 2 554 555 282 280 678  1  0  1.01
```

> In the next column you are clearly counting number of marital pairs ...

Nope, I'm not counting marital pairs anywhere. I'm counting:

- Individuals – living or dead, present or absent, square or not square, etc.
- Households as defined in the sources.
- In some cases neighborhoods within villages as defined in the sources.
- Camps or villages as defined in the sources.
- Regional populations containing several villages as defined in the sources.
- Also where available I include conceptual groups such as patrilineages, but most sources lack that kind of data so it is rare in the data base.

In the Summary I count each individual separately, classify them as living or dead, and report the results as Number of Living People and Total Number of Records. Then I count all of the other analyzable residential groupings from households through neighborhoods and camps / villages, to regional populations, and sum the numbers to get the figure that appears in the Number of Households, Villages, etc. column.

> This situation is in marked contrast to the Copper Eskimo picture where the
 > isolated nuclear family comprised most households or domiciled units. The group
 > composition charts of the Copper Eskimo are virtually indistinguishable from those
 > of Iglulik, so that this contrast in family organization is not shown on them.

In the Key to NU10 Copper 1922-23, we see the following data definitions:

10. ResGrp1 (1-2) Regional populations, Winter 1922-23
11. ResGrp2 (1-4) Villages in Regional Population #1, Spring 1923
12. ResGrp3 (1-32) Households in Regional Population #1, Spring 1923

Here we have censuses and genealogies for 2 large Regional Populations during the Winter of 1922-23 (ResGrp1). The Regional Populations serve as the bases for the two genealogical diagrams. I don't remember how Regional Population is defined, but if I need to know I can return to the sources.

Next we have censuses of 4 villages within the first Regional Population during Spring 1923 (ResGrp2).

Finally we have 32 households within the first Regional Population in the Spring (ResGrp3).

If we look at the NU09 Netsilik Key, we see the following:

10. ResGrp1 (1-6) Regional populations, Winter 1922-23

11. ResGrp2 (1-13) Villages, Spring 1923

12. ResGrp3 (1-80) Households, Spring 1923

This yields a total of 99 analyzable Netsilik residential units of some kind. The ethnographic details are specified in the data source. In the following excerpt from the NU09 Netsilik 1922-23 data file, the codes used for individuals in the Netsilik source appear in Var13. ID code used in source document. In this case the ID codes used in the source document indicate village (1.xx) and household (x.06) memberships and should correspond exactly with data in the source documents.

File.Rcrd.	1	2	3	4	5	6	7	8	9	10	11	12	13
NU09 0001	1	1	1	0	2	0	0	241	0	1	1	6	1.06
NU09 0002	2	1	1	0	1	14	252	0	0	1	1	6	1.06
NU09 0003	3	1	1	0	1	18	254	0	0	1	1	1	1.01
NU09 0004	4	1	1	0	1	21	255	0	0	1	2	12	2.02
NU09 0005	5	1	1	0	1	21	255	0	0	1	2	12	2.02

Conclusion.

One of my many objectives in compiling the GCBS Data Base was to develop an archive that is standardized in as many ways as is possible without forcing anything into it that would be uncomfortable there, and to provide direct linkages back to the original sources for anyone who wants more detail than the standardized format can accommodate. Standardization permits uniformity of processing and analysis, while links back to the sources ensure ethnographic richness and appropriateness.

Note that the last variable in the numerical data files is "ID Code used in source document". If you have the data source, you can use this link to go directly back to the context, clarify whatever I obscured and get all the fine points and nuances that I had to leave out. The numerical format does some things really well and some things poorly. I have tried to capitalize on its strengths and provide workarounds for dealing with its limitations.

I certainly agree with your concern about the numbers on the Statistical summary and I hope I have addressed all of them here. Perhaps my comments in the revised and expanded introduction to the Statistical Summary will be useful to you and others.

I hope to transfer all of my recently completed revisions to the server in California in a few days. As soon as all of the changes are in place and working properly, I'll let you know.

If you find additional problems, please contact me. I really value your expert knowledge of the Inuit datasets, and am confident that your comments on those datasets in particular and on the GCBS Database in general can stimulate me to make a lot of improvements.

Warmest regards,

Woody

2002-06-21 Damas to Denham

From: David Damas, Burlington, Ontario, CA
To: Woody Denham, Sharjah, UAE
Date: 21 June 2002
Subj: Central Arctic populations

Dear Woody,

I have tried several times this week to reach you by FAX but my contact reported that they succeeded in reaching only a phone response. I know that the two numbers are identical but in the case of the past two contacts there was an automatic switch over. At any rate I was merely letting you know that this packet would be reaching you via regular mail because of its bulk. Hope it gets to you without too much delay.

I'm glad that you sent a copy of your GCBS Statistical Summary. It clarifies some points and gives me the basis for some further comments and expansion of points made in my earlier letter.

First with regard to Number of Living People. Ok. You are including people residing outside the band's diagram. But it must be made clear that these missing people who are diagrammed within squares do not come close to representing all relatives for the people of the band or village. They are, rather, those people who are material to the structure of the band. Of course this presumes that kinship considerations were chief in recruiting members. June Helm elaborates on this assumption in material enclosed. The total genealogical network or marriage universe extends to about five hundred people for the Iglulik samples. For the Netsilik there are an additional 170 above those

featured in my diagrams. In the Copper Eskimo case the 206 are representative of over 800 in the total "tribe." Individuals will have siblings or other primary relatives scattered over these large areas. The same goes for Total Number of Records, i.e., living and dead. I have included only those deceased who somehow or other appear to create ties that help bind the groups together.

Turning to what I see as the most troublesome areas: Number of Households, Camps and Villages by Census. Here definitional problems intrude. [There are] two problems

(1) Use of term *households*.

Since you have said to base your conception on the description or apparent definition of your source, I have included a photocopy of my article on the Eskimo family. You will note a range in composition in the Central Arctic. You are in fact using the nuclear family as the unit you are calling household. I hope that my article will convince you that it is an unfortunate usage. Indeed in your comparative material, I think that you should stick to the nuclear family since in all societies (perhaps the Nayar exception) it occurs. I am glad that you are separating your households (actual nuclear families) from the local group count. As far as moving to the next stage of adding the two I still think that this is commingling apples with oranges and is a questionable and not useful step.

(2) What are you counting?

As for your denying that you are counting marital pairs how come my analysis of the 1921-22 Iglulingmiut group comes out to 48 of them which corresponds to your households? The count is not quite so close for the Nesilik sample where I get 86 as compared to your 91. Not very close for the Copper sample where I get 60 and you 39. If you are not counting nuclear families headed by marital pairs as your households what are you counting for them?

Turning then to the enclosed material:

a) Chapter IV from my Iglulingmiut Study. You remarked that you were distant from any library that would help and wanted more information so I copied the entire chapter.

b) "Characteristics of Central Eskimo Band Structure" and "The Problem of the Eskimo Family". These give comparative material for the Copper and Netsilik groups on band and household. I have also included two articles by Helm which develop her thoughts and methodology. Again I urge you to reconsider the marital pairing concept. While you write that you focus on the individual you also state that marital pairs can be abstracted. This is good. Also it might be well to note that Helm was directing her use of the pairings toward theoretical questions, the degree of internal integration by kinship and the male female ratio. I'm not quite clear on what problems you are interested in but it is clear that a number of research questions can be addressed when you get this project on the net.

c) The final photocopy series from Mathiassen illustrates rather nicely, I think, some of the points I have tried to make. First looking at the total of the Iglulik Eskimos for 1921-22, note that the Iglulingmiut are only a part of this total which comprises a better than 90% discrete marriage universe. Second in the text he indicates how people moved back and forth between the several major regions during their life time. I could add that scattering of primary kin could also be found spread over this area. Finally note how the census is broken up into nuclear family units .This runs counter to information on extended family household units which I derived from both informant recall and observation of the 1960-61 situation. To give some idea of how the structuring of the Mathiassen census distorts the family / household organization, compare the diagram of the Point Elizabeth (Itdjering) group headed by Aua with the segmentation into nuclear families of the same group on page 17. Again this is an artifact of census taking, not revealing of actual household or extended family organization. This was the practice in the other Fifth Thule censuses.

Dave

2002-08-30 Denham to Damas

To: David Damas, Burlington, Ontario, CA
From: Woody Denham, Sharjah, UAE
Date: 30 August 2002
Subj: Central Arctic populations

Dear David,

Thanks very much for your letter and the package of materials. I really do appreciate them.

With regard to Number of Living People and Total Number of Records, I'll upgrade the Key files for all of the Iglulik, Netsilik and Copper datasets to reflect your concerns.

With regard to the more troublesome problems you raised, I haven't had time to study your comments, the publications you sent and the datasets themselves. I understand your concerns and enormously appreciate your contributions toward solving the problems you have detected.

I have three new courses beginning on 1 October, and am in the throes of working on a paper with Dr. Douglas White at UC Irvine that is scheduled for presentation in New Orleans in November and publication early in 2003. So I'm not sure exactly when I'll be able to focus on the Central Arctic datasets. But I shall do it as soon as possible.

One of the great advantages of working with these materials on the web is that I can include your comments with the datasets so that anyone else who ever works with the data will know exactly what your concerns are. With your permission, I can scan your letters and edit them to eradicate the gremlin that haunts your electric typewriter. Then I can send copies to you for your approval before I include them on the web. Would you please allow me to do that? I think that would greatly increase the value of the datasets for years to come.

Regards, Woody

* * * * *

2002-09-28 Denham to Damas

To: David Damas, Burlington, Ontario, CA
From: Woody Denham, Sharjah, UAE
Date: 28 September 2002
Subj: Number of Living People and Total Population

Dear Dave,

I haven't had time to deal with all of your concerns, but I have made a start.

> *First with regard to Number of Living People. OK. You are including people residing outside the band's diagram. But it must be made clear that these missing people who are diagrammed with squares do not come close to representing all relatives for the people of the band or village. They are, rather, those people who are material to the structure of the band.*

Remember that in the genealogical diagrams:

- a) **squares** denote people (ordinarily small children) of **unknown sex**.
- b) a **dark symbol without an "X"** denotes a living person who resides outside these villages but serves as a **genealogical link** between villages.

As indicated in the following excerpts from the Key files for each dataset, I tried to make a very clear distinction between total population and diagrammed population.

> *The total genealogical network or marriage universe extends to about five hundred people for the Iglulik samples.*

The NU05 Iglulik 1921-22 Key file includes the following:

Iglulik Eskimo population in 1922: 504 (Mathiassen 1928:15)

Constituent Regional Populations	Number of People
▪ Aivilingmiut (southern)	165
▪ Iglulingmiut (central)	146
▪ Tanunermiut (northern)	193
TOTAL	504

When I say further down in that file that the number of living people included in the numerical dataset = 151, that means that the 151 people represented in the dataset come from the total of 504 that Mathiassen reported.

> *In the Copper Eskimo case the 206 are representative of over 800 in the total "tribe."*

The NU10 Copper1923 Key file includes the following line:

Copper Eskimo population in 1923: ca. 800 (Jenness 1922:277, Damas 1969:122)

When I say further down in that file that the number of living people included in the numerical dataset = 206, that means that the 206 people represented in the dataset come from the total of 800 that Jenness reported.

Therefore, in the Iglulik and Copper datasets, I think we are in complete agreement as indicated by what I have already included in the key files. But for clarity I have lightly edited these files so misunderstandings are less likely to occur.

> *For the Netsilik there are an additional 170 above those featured in my diagrams.*

There may be a real problem here, and I need your assistance to solve it. The NU09 Netsilik 1922-23 Key file includes the following line:

Netsilik Eskimo population in 1923: 259 (Rasmussen 1931:82)

Yet I say further down in that file that the number of living people included in the numerical dataset = 409 and the number of data records = 502.

From 20 years' distance and with no Netsilik documentation before me, I can't remember how I found 409 living people in a population that presumably consisted of only 259 people. The dataset and diagrams contain about 50 living links who were not part of the resident population, but that doesn't account for most of the discrepancy. Did I make an error in transcribing Rasmussen's figure: maybe I wrote 259 instead of 529? Or should it be 429 (259 + 170 = 429)? Or something else?

In any event, here my attempt to prevent confusion by stating the total population figure reported by the ethnographer seems instead to have created confusion. May I have your help in resolving this problem?

I hope to do some more work later today on the larger problems having to do with my use (or misuse) of "household", but will fax this to you now so further delays don't creep in.

Regards, Woody

* * * * *

2002-09-29 Denham to Damas

To: David Damas, Burlington, Ontario, CA
From: Woody Denham, Sharjah, UAE
Date: 29 September 2002
Subj: What I am counting; households; apples and oranges

Dear Dave,

First I deal strictly with what I am counting. Next I deal with the terminological problem associated with my use of "household". Finally I address your questions about my combining apples and oranges. I have not had time to read everything you sent, but all of the items are old friends.

Part 1. What am I counting?

> What are you counting? If you are not counting nuclear families headed by marital pairs as your households what are you counting?

> My analysis of the 1921-22 Iglulingmiut group comes out to 48 marital pairs which corresponds to your households.

> The count is not quite so close for the Nesilik sample where I get 86 as compared to your 91.

> Not very close for the Copper sample where I get 60 and you 39.

The bottom line is: I think we may be counting different things. I am printing three tables here so we can make sure we're talking about the same things.

Table 1. Excerpts from the NU05 Iglulik 1921-22 KEY file and DATA file

Table 2. Excerpts from the NU09 Netsilik 1922-23 KEY file and DATA file

Table 3. Excerpts from the NU10 Copper 1922-23 KEY file and DATA file

Tables 1-3 have three parts.

- The top part (KEY) contains a detailed description of the variables (and their values) that appear in the numerical data files
- The middle part contains a brief tabular summary of residential group sizes for each dataset.
- The bottom part contains the column header record and the first few data records from the numerical data files.

Please see my comments following each table.

Sorry this is so long. You ask interesting and difficult questions.

1. Excerpts from NU05 Iglulalik 1921-22 Key

KEY

1. ID number
2. Life status 1=alive, 0=dead
3. Sex: 1=male, 2=female
4. Age: not available
5. Current marital status: 0=not applicable, 1=never married, 2=married, 4=widowed
6. Father's ID
7. Mother's ID
8. 1st spouses' ID
9. 2nd spouses' ID
10. 3rd spouses' ID
11. ResGrp1 (1-5) Villages, Winter 1921-22
12. ResGrp2 (1-48) Households, Winter 1921-22
13. ID code used in source document.

Variable number	Variable name and values	Number of people	Mean size of residential units
11	Iglulingmiut Villages, Winter 1921-22		
	V# 1	26	
	V# 2	10	
	V# 3	17	
	V# 4	74	
	V# 5	16	
	Total	143	28.6
12	Iglulingmiut Households, Winter 1921-22		
	HH# 1-48	143	3.0

Table 3. Residential group sizes.

File.Rcrd	1	2	3	4	5	6	7	8	9	10	11	12	13
NU05 0001	1	1	1	0	2	4	105	103	0	0	2	11	2.02
NU05 0002	2	1	1	0	1	4	105	0	0	0	2	10	2.01
NU05 0003	3	1	1	0	2	0	0	104	0	0	2	12	2.03
NU05 0004	4	1	1	0	2	0	0	105	0	0	2	10	2.01
NU05 0005	5	1	1	0	1	7	112	0	0	0	1	8	1.08
NU05 0006	6	1	1	0	1	14	115	0	0	0	1	4	1.04
NU05 0007	7	1	1	0	2	18	119	112	0	0	1	8	1.08
NU05 0008	8	1	1	0	1	18	119	0	0	0	1	7	1.07
NU05 0009	9	1	1	0	1	18	119	0	0	0	1	7	1.07
NU05 0010	10	1	1	0	1	19	120	0	0	0	0	0	0
NU05 0011	11	1	1	0	1	19	120	0	0	0	0	0	0

Excerpt from data file

Discussion

> *My analysis of the 1921-22 Iglulingmiut group comes out to 48 marital pairs which corresponds to your households.*

Var13. In the top section of the table (Key), we see that Var13 in the data file holds the ID code used in the source document.

13. ID code used in source document.

So if we look at the bottom section of the table (Data), we see that Var13 at the extreme right end of each data record contains the ID code used in the source document.

All of the Central Inuit data files use the xx.yy format that we see in Var13 of the NU05 Iglulik 1921-22 data file, where xx = village number and yy = something that I translated as “household” in my generic, generalizing, worldwide terminology. (I’ll consider the appropriateness of that term later. Here I’m concerned only with defining it in the context of the Inuit datasets.)

In theory, if we lay a hardcopy of the data file on the table beside the source document from which it was extracted, we can see where each person in the data file came from. In practice the match between the values in Var13 and the source codes may be somewhat indirect due to several factors including the various ways in which source codes were assigned by ethnographers (sometimes none were used, sometimes only a few were assigned to key people, sometimes the same code was used over and over for different people, etc.), and my decision to work from two or more sources whenever possible. In the best case, the code in Var13 tells you exactly who you’re looking at in the original data source; in the worst case, you may have to be pretty creative to get from the one to the other. The Inuit datasets are problematic here.

Village codes: In NU05, I assigned village numbers (1-5). If you compare my diagrams for NU05 with the diagrams in Igluligmiut Kinship and Local Groupings, p.60-62, you’ll see the following which I have added to the Key file (a single NU05 diagram shows Villages 3+4 as on p.62 in the document):

Village 1 = Itibjiriao	(p.60)
Village 2 = Amitsuq	(p.60)
Village 3 = Iglulik	(p.62)
Village 4 = Pingirqalik	(p.62)
Village 5 = Manirtuq	(p.60)

Household codes: I can’t remember exactly how I derived the “household” codes for the various Central Inuit datasets and I lack the data I would need to reconstruct the algorithm. With regard to NU05 Iglulik, since both of us found 48 units, I infer that both of us used the same procedure to identify those units. In this case, the unit of analysis seems to be the nuclear family or marital pair plus co-resident children, probably

because, as you note, those were the units of analysis in the original source documents from the 5th Thule Expedition. The same applies to the Netsilik and Copper data. In other words, I always tried to take the raw data as I found it and standardize its format without changing it to fit any other interpretive schemes.

If my coding here and elsewhere accurately reflects the original coding scheme from 5th Thule but you feel it should include a different set of codes based on extended family units that were not demarcated in the original data, we can insert another variable for Extended Families whose numerical values would represent these larger units in which the people lived. In other words, the two coding schemes for “household-like” residential units would complement each other.

Var11. If we look next at the top section of the table (Key), we see that Var11 in the data file holds the code for the village in which each person resided at the time of the census in question.

11. ResGrp1 (1-5) Villages, Winter 1921-22

So if we look at the bottom section of the table (Data), we see that Var11 contains codes for villages. The source codes in Var13 use village numbers 1-5, so that is what I used in Var11. The first data record has a source code (Var13) = 2.02, so the village code (Var11) = 2. They are the same. In general, then, for this dataset, Var13 contains source codes for 5 villages and Var11 uses a corresponding set of 5 village codes (1-5).

When you look at the middle of the table, you see the following tabulation for Var11, showing five villages (1-5), their individual sizes, their total population (143), and their mean size (28.6).

11	Iglulingmiut Villages,		
	Winter 1921-22		
	V#	1	26
	V#	2	10
	V#	3	17
	V#	4	74
	V#	5	16
	Total	5	143
			28.6

So my data in Var11 came directly from the village number in Var.13.

Var12. Now turn to Var12.

12. ResGrp2 (1-48) Households, Winter 1921-22

This variable is constructed exactly the same way that Var11 was constructed, with one minor change. In the source documents, the “households” in each village were coded from 1 to the maximum number in the village. (Compare your diagrams for Itibjiriaq and Manirtuq, both of which contain 1,2,3,4,5). In the NU05 dataset, the “household” codes

that correspond to codes used in the source documents appear in the left column below.

Source code	Recoded Var12
1.01-1.09	1 through 9
2.01-2.03	10 through 12
3.01-3.06	13 through 18
4.01-4.25	19 through 43
5.01-5.05	44 through 48

Throughout the GCBS Data Base, I wanted to use a single uninterrupted series of numbers for “households” instead of several short series each of which began again at 1. So I recoded what appeared in the source codes and used the recoded values in Var12 as shown in the right column above. In this dataset, Village 1 had 9 “households” numbered 1-9, and I retained those numbers in Var12. Village 2 had 3 “households” numbered 1-3, so I converted them to 10-12, and so on. There were 48 such “households” in the source documents, so Var12 contains values 1-48.

In the middle of the table above, you see the following tabulation for Var12, showing 48 households, their total population (143), and their mean size (3.0).

```
12      Iglulingmiut Households,
        Winter 1921-22
        HH#  1-48  143      3.0
```

As in the case of Var11, my data in Var12 came directly from Var.13.

Since both of us concluded that the NU05 Iglulik 1921-22 dataset contains 48 of these units (whatever we may call them), it seems that we are counting the same things. That is good. Later I’ll return to the question of what I should call them. For now, I’m relieved that both of us found the same number = 48.

Statistical Summary Now look at the statistical summary for NU05 Iglulik:

Society code	Society name	Date of ethnographic present	Number of living people	Total number of records	Number of households, camps, villages X censuses	Diachronic data?
NU05	Iglulik	1921-22	151	178	53	Yes

Number of living people =151 even though the total number of people in the 5 villages = 143. The difference lies in the nonresident linking individuals who were mentioned in the source document and included in the data file, but were omitted from the tabulation since they were not residents of these villages. They are included for genealogical purposes, but excluded for census purposes.

Total number of records = 178, including 153 living people and 25 deceased people who were explicitly included in the original diagrams, explicitly mentioned in the original documentation, or unambiguously implied in the original documentation.

Number of households, camps, villages X censuses = 53, which is the sum of the 5 villages and 48 households. Certainly this mixes apples and oranges, and I'll return to discuss it later. Here I simply want to indicate which apples and which oranges I was counting.

2. Excerpts from NU09 Netsilik 1922-23 Key

KEY

1. ID number
2. Life status 1=alive, 0=dead
3. Sex: 1=male, 2=female
4. Age: not available
5. Current marital status: 0=not applicable, 1=never married, 2=married, 4=widowed
6. Father's ID
7. Mother's ID
8. 1st spouses' ID
9. 2nd spouses' ID
10. ResGrp1 (1-6) Regional populations, Winter 1922-23
11. ResGrp2 (1-13) Villages, Spring 1923
12. ResGrp3 (1-80) Households, Spring 1923
13. ID code used in source document.

```

-----
Variable  Variable      Number      Mean size of
number   name and      of people   residential
         values                                     units

10       Regional populations (6)
         Winter 1922-23
         RP#  1-6   363      61

11       Villages (13)
         Spring 1923
         V#   1-13  363      28

12       Households (80)
         Spring 1923
         HH#  1-80  363      4.5
  
```

Table 3. Residential group sizes.

```

-----
File.Rcrd....1.2.3..4.5...6...7...8....9.10.11.12..13
NU09 0001    1 1 1  0 2   0  0 241    0 1  1  6  1.06
NU09 0002    2 1 1  0 1  14 252   0   0 1  1  6  1.06
NU09 0003    3 1 1  0 1  18 254   0   0 1  1  1  1.01
NU09 0004    4 1 1  0 1  21 255   0   0 1  2 12  2.02
NU09 0005    5 1 1  0 1  21 255   0   0 1  2 12  2.02
NU09 0006    6 1 1  0 2  21 255 246   0 1  2 12  2.02
NU09 0007    7 1 1  0 1  22 256   0   0 1  2 11  2.01
NU09 0008    8 1 1  0 1  22 256   0   0 1  2 11  2.01
NU09 0009    9 1 1  0 1  22 256   0   0 1  2 11  2.01
NU09 0010   10 1 1  0 1  23 257   0   0 1  1 10  1.10
  
```

Excerpt from data file

Discussion

> *The count is not quite so close for the Nesilik sample where I get 86 as compared to your 91.*

In extracting the NU09 Netsilik 1922-23 dataset from the source documents, I used the same procedures that I used and described in detail above for the NU05 Iglulik 1921-22 dataset. However, the NU09 Netsilik source showed THREE levels of organization (Region, Village, Household) rather than the two levels shown for NU05 dataset (Village, Household). Therefore, I used three variables (Var10,11,12) for analyzable data, and Var13 again for the source code.

- 10. ResGrp1 (1-6) Regional populations, Winter 1922-23
- 11. ResGrp2 (1-13) Villages, Spring 1923
- 12. ResGrp3 (1-80) Households, Spring 1923
- 13. ID code used in source document.

Var13. From here I can't reconstruct the procedure I used to generate the household portion of the code in Var13 for I don't have the source documents for NU09 Netsilik. If I misinterpreted the coding in the source documents so that the values I entered into Var13 do not correspond to that which was intended in the document, then Var13 may have a problem and everything that follows from Var13 may have a problem too. This problem lies at the core of what I am counting and calling "households". A careful examination of the diagrams and numerical data indeed suggests that the unit of analysis here as in the Iglulik case was indeed the nuclear family or marital pair plus co-resident children.

Var10. Notice that the Var10 Regional Population census was for Winter 1922-23, whereas the Var11 Villages and Var12 Households censuses were for Spring 1923. This means of course that the number of people in the Winter census need not have matched the number in the Spring census if some people had moved into or out of the area between the two counts. It also means that Var11 and Var12 were not simply a different way of organizing the data in Var10; no doubt the Winter and Spring social configuration of people on the ground had a lot in common, but in this dataset the relationship between Regional Populations and Villages is NOT logically equivalent to the relationship between Spring Villages and Spring Households, where households were constituent parts of the villages.

The middle section of the table shows the same kind of tabulation I did for the NU05 Iglulik dataset, but it has three levels:

10	Regional populations (6)			
	Winter 1922-23			
	RP#	1-6	363	61
11	Villages (13)			
	Spring 1923			
	V#	1-13	363	28

12 Households (80)
 Spring 1923
 HH# 1-80 363 4.5

As in the NU05 Iglulik dataset, my data in Var11 and Var12 came directly from Var.13. The data for Var10 was not embedded in the source codes in Var13, so they came from the documents themselves.

The Statistical Summary shows 409 living people (363 residents and 46 nonresidents) and 502 total records including 93 deceased ancestors. Also it shows 99 residential units: 6 regional +13 villages + 80 households = 99 units.

Society code	Society name	Date of ethnographic present	Number of living people	Total number of records	Number of households, camps, villages X censuses	Diachronic data?
NU09	Netsilik	1922-23	409	502	99	No

At the top of this page you said, “I get 86 as compared to your 91”. But I don’t get 91. I get 80 “households” and 99 “apples-and-oranges residential units”, but I don’t get 91 of anything. To me, that says we’re counting different things.

3. Excerpts from NU10 Copper 1922-23 Key

KEY

1. ID number
2. Life status 1=alive, 0=dead
3. Sex: 1=male, 2=female
4. Age: not available
5. Current marital status: 0=not applicable, 1=never married, 2=married, 4=widowed
6. Father's ID
7. Mother's ID
8. 1st spouses' ID
9. 2nd spouses' ID
10. ResGrp1 (1-2) Regional populations, Winter 1922-23
11. ResGrp2 (1-4) Villages in Regional Population #1, Spring 1923
12. ResGrp3 (1-32) Households in Regional Population #1, Spring 1923
13. ID code used in source document.

Variable number	Variable name and values	Number of people	Mean size of residential units
10	Regional populations (2) Winter 1922-23		
	RP# 1	107	
	RP# 2	99	
	Total 2	206	103
11	Villages (4) Spring 1923		
	V# 1	15	
	V# 2	31	
	V# 3	33	
	V# 4	28	
	Total 4	107	27
12	Households (32) Spring 1923		
	HH# 1-32	107	3.3

Table 3. Residential group sizes.

File.Rcrd.	1	2	3	4	5	6	7	8	9	10	11	12	13
NU10 0001	2	1	1	0	1	28	142	0	0	1	3	14	3.14
NU10 0002	3	1	1	0	1	28	142	0	0	1	3	14	3.14
NU10 0003	4	1	1	0	1	28	142	0	0	1	3	14	3.14
NU10 0004	5	1	1	0	1	29	143	0	0	1	3	18	3.18
NU10 0005	6	1	1	0	1	56	145	0	0	1	2	8	2.08
NU10 0006	7	1	1	0	1	33	146	0	0	1	2	10	2.10
NU10 0007	8	1	1	0	1	33	146	0	0	1	2	10	2.10
NU10 0008	9	1	1	0	1	35	147	0	0	1	2	7	2.07

NU10 0009 10 1 1 0 1 36 141 0 0 1 3 16 3.16

Excerpt from data file

Discussion

> *Not very close for the Copper sample where I get 60 and you 39.*

The procedures for extracting and coding the data are the same here as described above.

10. ResGrp1 (1-2) Regional populations, Winter 1922-23
11. ResGrp2 (1-4) Villages in Regional Population #1, Spring 1923
12. ResGrp3 (1-32) Households in Regional Population #1, Spring 1923
13. ID code used in source document.

But here another complication intrudes. Notice that Var10 pertains to TWO Regional Populations in Winter 1922-23, whereas Var11 and Var12 pertain to Villages and Households belonging ONLY to Regional Population 1, and they were censused not in the Winter but in the Spring. The tabular data below shows the numbers for each group. RP1 had 107 members and RP2 had 99 for a total of 206, which is the number that appears as the number of living people in the Statistical Summary. The total number of residential unites is the sum: 2 Winter Regional Populations + 4 Spring Villages + 32 Households within the Spring Villages = 38 “apples-and-oranges residential units”.

```

10      Regional populations (2)
        Winter 1922-23
          RP#  1      107
          RP#  2      99
        Total  2      206      103

11      RP#1 Villages (4)
        Spring 1923
          V#   1      15
          V#   2      31
          V#   3      33
          V#   4      28
        Total  4      107      27

12      RP#1 Households (32)
        Spring 1923
          HH#  1-32  107      3.3
  
```

Society code	Society name	Date of ethnographic present	Number of living people	Total number of records	Number of households, camps, villages X censuses	Diachronic data?
NU10	Copper	1922-23	206	272	38	No

As in the NU05 Iglulik and NU09 Netsilik dataset, my data in Var11 and Var12 came directly from the source code in Var.13. Again the data for Var10 was not embedded in the source codes in Var13, so they came from the documents themselves.

At the top of this page you said, “I get 60 and you 39”. But I don’t get 39. I get 32 “households” and 38 “apples-and-oranges residential units”, but I don’t get 39 of anything. Again, this says to me that we’re counting different things.

Since I have no access to my source data, I don’t know how we came out with such a large difference. Might it have something to do with the fact that the Var11 villages and Var12 households included in the dataset come ONLY from what I have called Regional Population #1? If your count comes from both Regional Populations and mine comes only from RP1, that would account for the difference. It doesn’t seem possible that the difference could come from my counting nuclear families or marital pairs and your counting extended family households – that should make my number bigger than yours rather than smaller.

Part 2. What should I call it?

I fully understand your concern about “What’s in a name?” If I label something incorrectly, people get confused. But what is the best label?

I have used a rough and ready hierarchy: Regional Population > Villages/Camps > Neighborhoods (optional multi-residence/multi-household segments of villages / camps) > Households > People, where “>” means something like “is comprised of”.

I need – but haven’t found - a good term for a group that lives together, probably consists of people who are related to each other in some way and occupies one residential structure, is smaller than the next level of complexity (neighborhood, village/camp, troop, etc.) and is as atheoretical and culture-free as possible.

Part of the problem is theoretical. Household compositions and band compositions (bands, tribes, dialect groups) have been the focal points of often heated theoretical debates for decades. Perhaps the safest course would be to avoid all terms that have ever found their way into those debates. All of them have been somehow “tainted” and every one of them is likely to offend somebody.

Part of the problem is ethnographic. Aborigines, Iglulik, San, Ainu and Mbuti do not organize themselves in precisely the same way, so the terminology I use must fit the ethnographic reality of the group to whom I apply it. I understand and accept that point, but in reply I turn to the next point.

Part of the problem is my attempt to construct a “global” database that contains an enormous diversity of cultures, and to do it in such a way that all datasets can be analyzed in the same way, yielding meaningful statements about ALL of the societies represented here despite their ethnographic diversity.

Upon examining the GCBD Database Manual, I see that I didn’t pay much attention to defining ordinary language terms such as “household”. Rather I attempted to use that term in its broadest and most generic and comprehensive sense, with the (excessively

implicit) understanding that an ethnographer working with Aboriginal data would understand the term as it applies to Australia or to Central Australia, while an Inuit specialist would interpret it accordingly when working with Central Inuit data. Clearly this is a cavalier approach and I must do something to fix it. But what?

I won't set up the database in such a way that it contains major theoretical arguments built directly into its very infrastructure. It isn't possible to make GCBS or anything like it totally devoid of theory, but it is entirely legitimate to attempt to make it as atheoretical as I possibly can. Subverting it with theoretically load ideas would defeat the purpose of analyzing the data to discover what's in there. If what's in there derives from my theory rather than from whatever the natives were doing, the exercise is pointless.

Likewise, I won't set up the database in such a way that each dataset must be dealt with in its own unique set of concepts, its own ethnographic vocabulary. That defeats the purpose of making it "global" or "universal". I want people to be able to analyze it for what it says about the human species as a whole, not just what it says about their own culture or culture area.

So the terminology I use must be as atheoretical as possible and as culture-free as possible. That does not mean it can be incorrect or misleading or infuriating, but it must not bias the data in any way. That's hard to do.

For example, let's consider "marital pairs". It might work with Iglulik, but I suspect it would have a hard time accommodating polygynous unions among Central Australian Aborigines where a man has multiple wives living with him in his own residence. "Household" may be a sloppy way to characterize such a residential unit, but "marital pair" simply won't work. (In the NU05 Iglulik dataset I see only two cases of polygyny, and I coded both cases to show the husband and his two wives in one and the same "household" – but not the same "marital pair".) Perhaps "nuclear family" is a better way to say it, but when a polygynous nuclear family occupies a residential unit as part of an extended family with grandma and a couple of cousins and their wives living in it, "household" seems to have a better intuitive meaning than either "marital pair" or "nuclear family".

I don't reject marital pair or nuclear family, but I remain skeptical of their general utility. I'm willing to consider anything that it is NOT culture-specific or theoretically loaded, and is general (even vague) enough to do the job for all hunter-gatherers.

Part 3. Apples and oranges

The column in the Statistical Summary called "Number of households, camps, villages X censuses" is included in the summary ONLY as a very simple way of indicating something rather nebulous about the richness of a particular dataset. It does not in any sense say anything about the people to whom the set pertains nor does it have any theoretical significance. As a simple "index of complexity" I think it is informative. Furthermore, I have been unable to create anything else that serves as such an index. But if I had a better one, I would use it.

Perhaps I can further edit the text at the top of the Statistical Summary file to make it clear that this is a ballpark estimate of complexity to which no significance should be attached. But if it presents a real problem and needs to go, I could consider deleting it without losing anything of great value.

Regards, Woody