

*И.М.Гарскова (Москва)*

**"От просопографии к статистике":  
МЕТОДИКА АНАЛИЗА БАЗ ДАННЫХ ПО ИСТОЧНИКАМ,  
СОДЕРЖАЩИМ ДИНАМИЧЕСКУЮ ИНФОРМАЦИЮ**

Анализ динамических рядов на основе довольно сложных статистических моделей пока довольно редко встречается в работах отечественных историков (исключением в практике применения динамических моделей, пожалуй, является область экономической истории).

В данной работе предлагается подход к анализу информации баз данных, которые содержат динамические сведения. Такие базы данных обычно создаются при работе с источниками, отражающими "жизненные пути" или "коллективные биографии" различных социаль-

ных, профессиональных или иных групп людей, и носят название просопографических баз данных. Типичным методическим подходом в этом случае является создание многофайловой базы данных, один из файлов которой содержит основные статические сведения и связывается с одним или многими файлами, содержащими динамическую информацию и связанными с основным файлом ключевым полем - идентификатором.

В качестве источников могут быть использованы биографические сведения о политических деятелях (например, членах парламента), извлеченные из специальных справочных изданий, энциклопедий и т. п., особенно если эти персоналии многократно входили, в том числе и с перерывами, в изучаемую совокупность, и с течением времени их профессиональный, социальный или иной статус менялся. Другими примерами могут быть данные личных дел контингента рабочих некоторой фирмы, социально-демографические данные о населении определенного региона с учетом миграций и т. п.

Вопрос конструирования базы данных является основным для подхода, который носит название "от статистики к просопографии" и уделяет основное внимание проблемам формирования "метаисточников", стандартизации имен и связыванию записей (record-linkage) о различных персоналиях, извлекаемых из целого ряда исходных, традиционных источников. Для предлагаемого здесь подхода, который можно назвать "от просопографии к статистике", центр тяжести лежит в другом: данные представляют наибольший аналитический интерес именно в агрегированном виде, как основа определенных динамических группировок, тогда как прослеживание "жизненных путей" отдельных персоналий не является первоочередной задачей - здесь наиболее интересны именно статистические результаты, обобщенные на уровне определенных групп персоналий в изучаемой совокупности, их структуры и динамики.

Просопографическая база данных соединяет сведения, содержащие как статичную (или уникальную) информацию, так и динамические сведения о каждой персоналии. Статичная информация обычно включает данные об имени, национальности, возрасте, месте рождения и других показателях, которые фиксируются однократно. Динамическая же информация включает данные, которые фиксировались многократно, при каждом их изменении: это изменения социального или семейного положения, перемещения с одного места работы на другое и т. д. Таким образом, наиболее явной особенностью просопографической базы данных является ее динамический, "трехмерный" характер, поскольку она содержит данные о каждом человеке в различные моменты времени, позволяя изучать изменения, касающиеся

не только отдельных персоналий, но и основных социально-профессиональных, возрастных и иных групп в структуре изучаемого контингента в различные периоды времени. Другой особенностью просопографической информации является ее "дискретный" характер: многие персоналии могут не принадлежать к анализируемой группе в течение всего изучаемого периода.

При решении задачи изучения структуры всей совокупности персоналий в ее динамике исследователю приходится учитывать обе особенности источника: каждый конкретный человек мог входить в ту или иную социальную, профессиональную, возрастную, национальную группу в зависимости не только от его меняющихся во времени характеристик, но и от того, присутствовал или отсутствовал он в каждый конкретный момент времени в "своей" группе.

Следовательно, изучая динамику состава анализируемой группы персоналий, мы в первую очередь сталкиваемся с проблемой построения динамических рядов по всем основным показателям для изучаемого периода. Построение таких динамических рядов опирается на возможность извлекать из базы данных списки персоналий, присутствующих в данной группе в каждый момент времени, и подсчитывать на основе этих списков числовые значения количественных показателей и доли встречаемости отдельных категорий качественных показателей (например, средний возраст или долю лиц с высшим образованием).

Для решения этой задачи автором создан меню-управляемый пакет программ ATiSeP (Aggregated Time Series on Prosopography - Агрегированные Динамические Ряды Просопографических данных) на языке dBASE IV, предназначенный для извлечения информации из многофайловой базы данных и построения динамических рядов полей этой базы для каждого указанного пользователем периода времени. Система работает с несколькими файлами базы данных: основным файлом, содержащим статические сведения о персоналиях (число записей в этом файле равно числу персоналий в базе данных), справочным файлом, содержащим по крайней мере одну запись для каждой персоналии о датах каждого "появления" в данной группе, и несколькими дополнительными файлами, содержащими сведения об изменениях различных динамических показателей и их датах по всем персоналиям, о которых такие сведения имеются.

Пакет состоит из четырех модулей: 1) модуля, формирующего структуры указанных файлов; 2) модуля, формирующего динамические ряды годовых или месячных данных; 3) модуля, формирующего вспомогательные (выборочные) базы данных, содержащие сведения о тех персоналиях, которые присутствуют в данном периоде;

4) модуля, экспортирующего эти выборочные базы данных в графический или статистический пакет для дальнейшего анализа информации во всей ее полноте.