

Использование больших данных в официальной статистике – новый раздел сайта Статкомитета СНГ

Новые технологии, широкое использование цифровых устройств, повсеместная генерация цифровой информации делает доступной в реальном времени информацию из различных источников, таких как GPS навигаторы, камеры видеонаблюдения, спутниковые данные, мобильные телефоны, электронная торговля, банковские карты, социальные сети, интернет запросы в поисковиках, электронные сообщения, карты местности пользователя и др. Если все подобные данные накапливать для дальнейшей обработки, то их суммарный объем будет измеряться десятками и сотнями петабайт.

Совокупность объемных и неструктурированных данных из всех таких источников принято обозначать термином «большие данные».

БОЛЬШИЕ ДАННЫЕ – это изначально не систематизированные данные о социально-экономическом развитии и окружающей среде, генерируемые на основе информационно-коммуникационных технологий вне официальной статистики.

Большие данные характеризуются такими признаками, как большой объем информации, высокая скорость изменения информации, разнообразие и разнородность данных. Эти характеристики определяют стратегическую ценность больших данных для официальной статистики:

- большой объем данных может способствовать получению более детализированных сведений по конкретным направлениям статистики;
- высокая скорость изменения информации способствует увеличению частоты статистических оценок;
- разнообразие больших данных может открыть возможности для производства статистики в новых областях (например, дистанционного зондирования и географических информационных систем), получения информации о не наблюдаемых официальной статистикой сферах;
- разнородность происхождения больших данных может способствовать многоаспектности измерения и, тем самым, повышению надежности статистических оценок.

Большие данные присутствуют едва ли не во всех сферах деятельности человека, в первую очередь – научно-исследовательской и коммерческой.

Проблема взаимодействия официальной статистики и больших данных уже несколько лет серьезно рассматривается в мире. В числе основных игроков (ведущих мировых центров, рассматривающих эту тему) стоит выделить следующие.

Статистическая комиссия ООН – высший мировой форум официальной статистики – на своей 45-й сессии (март 2014 года) начала подробное рассмотрение темы Больших данных.

В марте 2015 года на 46-й сессии Статистической Комиссии ООН был представлен доклад, в котором было дано определение «больших данных», приведены результаты анкетирования стран по организационным аспектам использования больших данных, существующим и будущим проектам их использования.

ГЛОБАЛЬНАЯ РАБОЧАЯ ГРУППА ПО БОЛЬШИМ ДАННЫМ В ОФИЦИАЛЬНОЙ СТАТИСТИКЕ (Global Working Group on BigData for official statistics) – образована решением 45-й сессии Статистической комиссии ООН для разработки руководства по вопросам больших данных. 31 октября 2014 года состоялась первая встреча Глобальной рабочей группы, на которой был представлен **Доклад Глобальной рабочей группы по вопросам использования больших данных для подготовки официальной статистики.**

ПРОЕКТ ООН «ГЛОБАЛЬНЫЙ ПУЛЬС» (UN Global Pulse) – созданная по инициативе Генерального секретаря ООН сеть инновационных лабораторий, координирующих исследования применения больших данных в магистральных направлениях деятельности ООН: климат и устойчивость, защита конфиденциальности данных, экономическое благополучие, пища и сельское хозяйство, гендерные вопросы, гуманитарные вопросы, здоровье общества, цели устойчивого развития после 2015 года, оценка в режиме реального времени.

ВСЕМИРНЫЕ КОНФЕРЕНЦИИ ПО ИСПОЛЬЗОВАНИЮ БОЛЬШИХ ДАННЫХ В ОФИЦИАЛЬНОЙ СТАТИСТИКЕ отражают признание необходимости выработки общих подходов и рассмотрения проблем больших данных.

Исходя из актуальности данной темы, Статкомитет СНГ разместил на своем сайте новый раздел «**Использование больших данных в официальной статистике**». На нем размещены основные материалы по данной тематике.