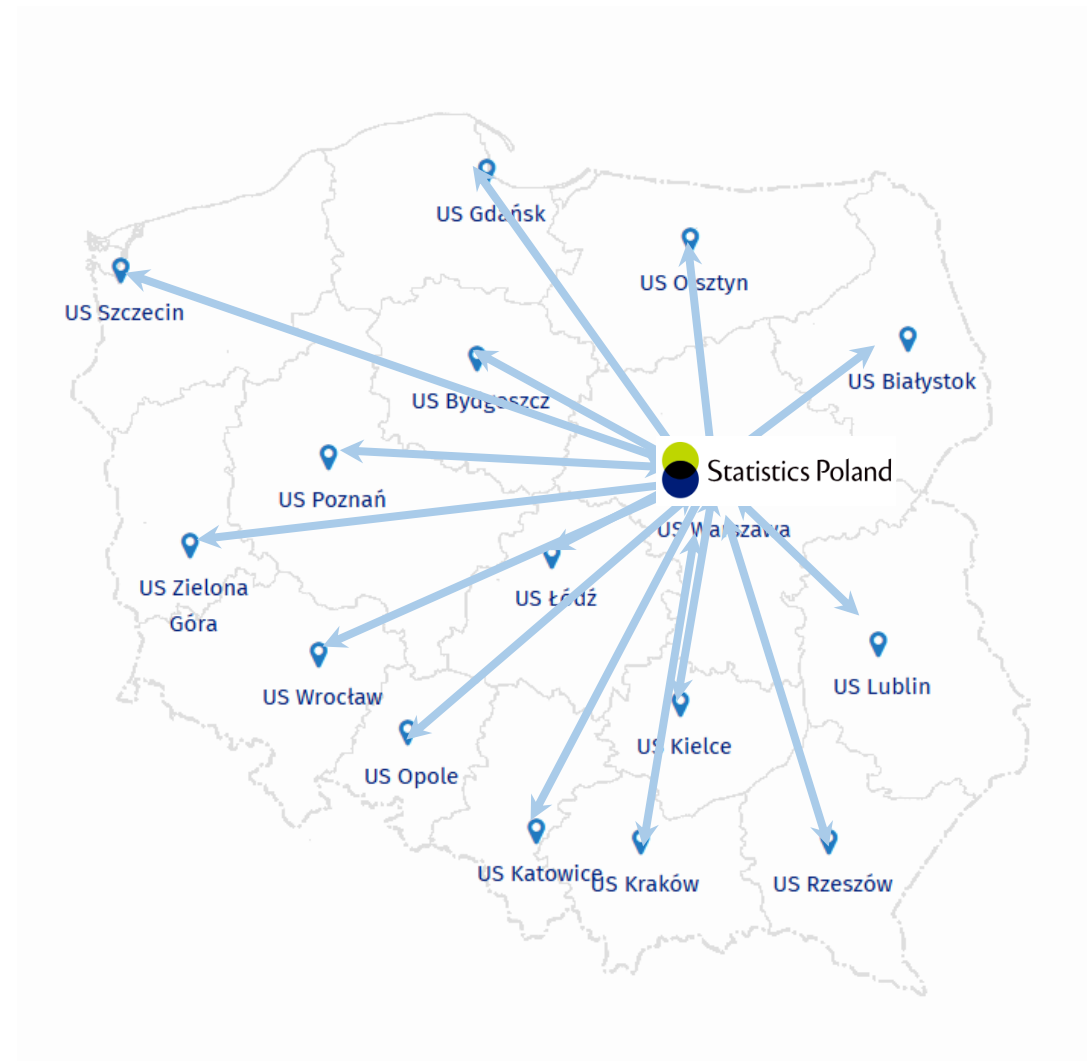


Качество данных и рост их ценности в современном мире

Опыт Статистики Польши — с точки зрения цепочки создания ценности данных

Доминика Рогалиньска
Вице-Президент

Организационная структура Статистики Польши



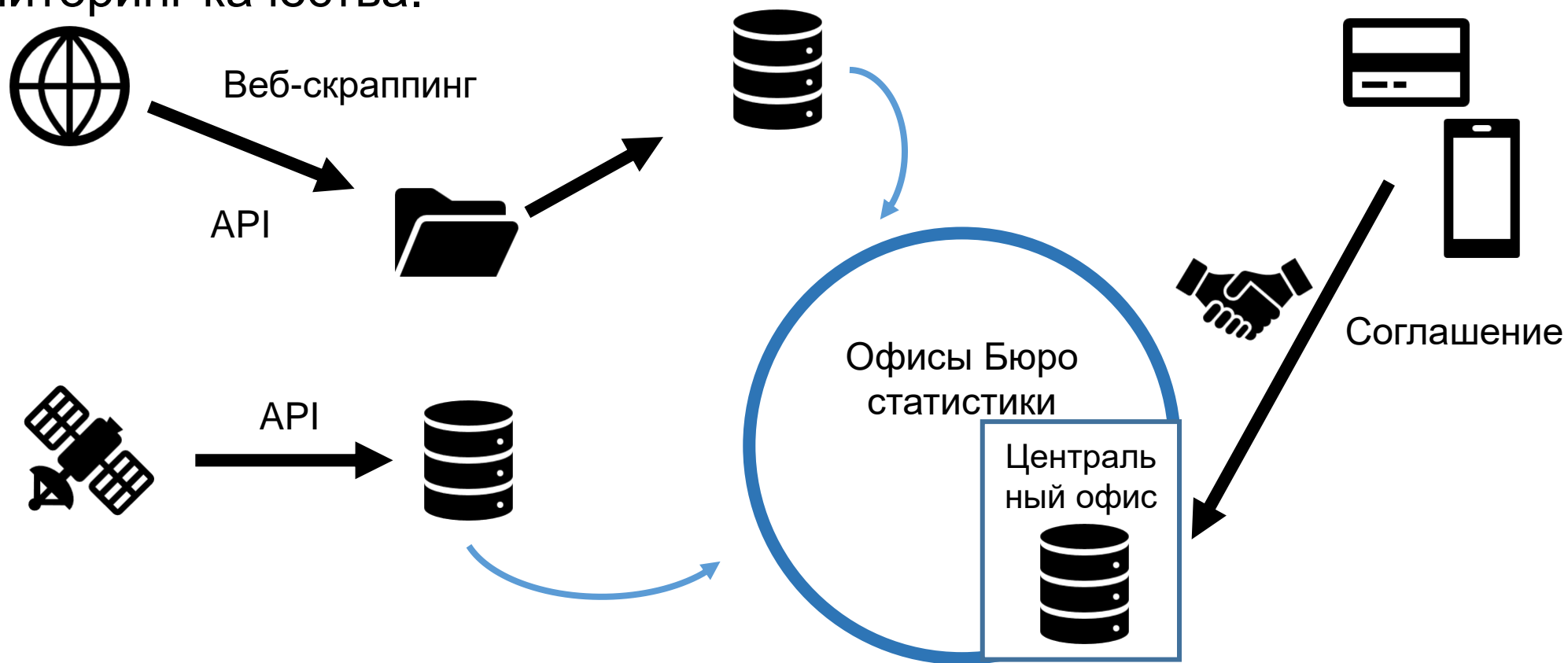
16 региональных
статистических
управлений

и

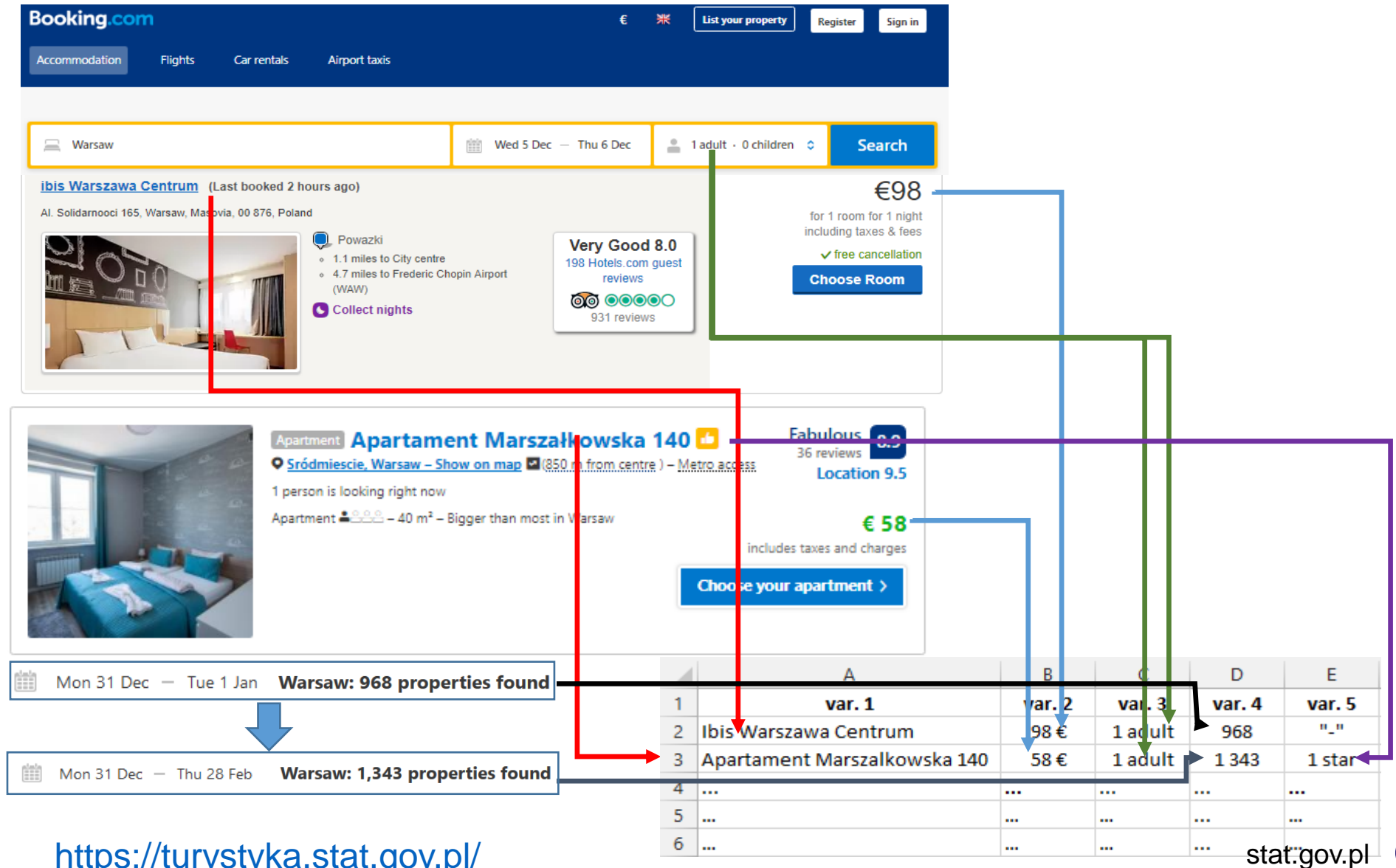
1 центральное
статистическое
управление в Варшаве

Обеспечение качества данных на каждом этапе цепочки создания ценности данных

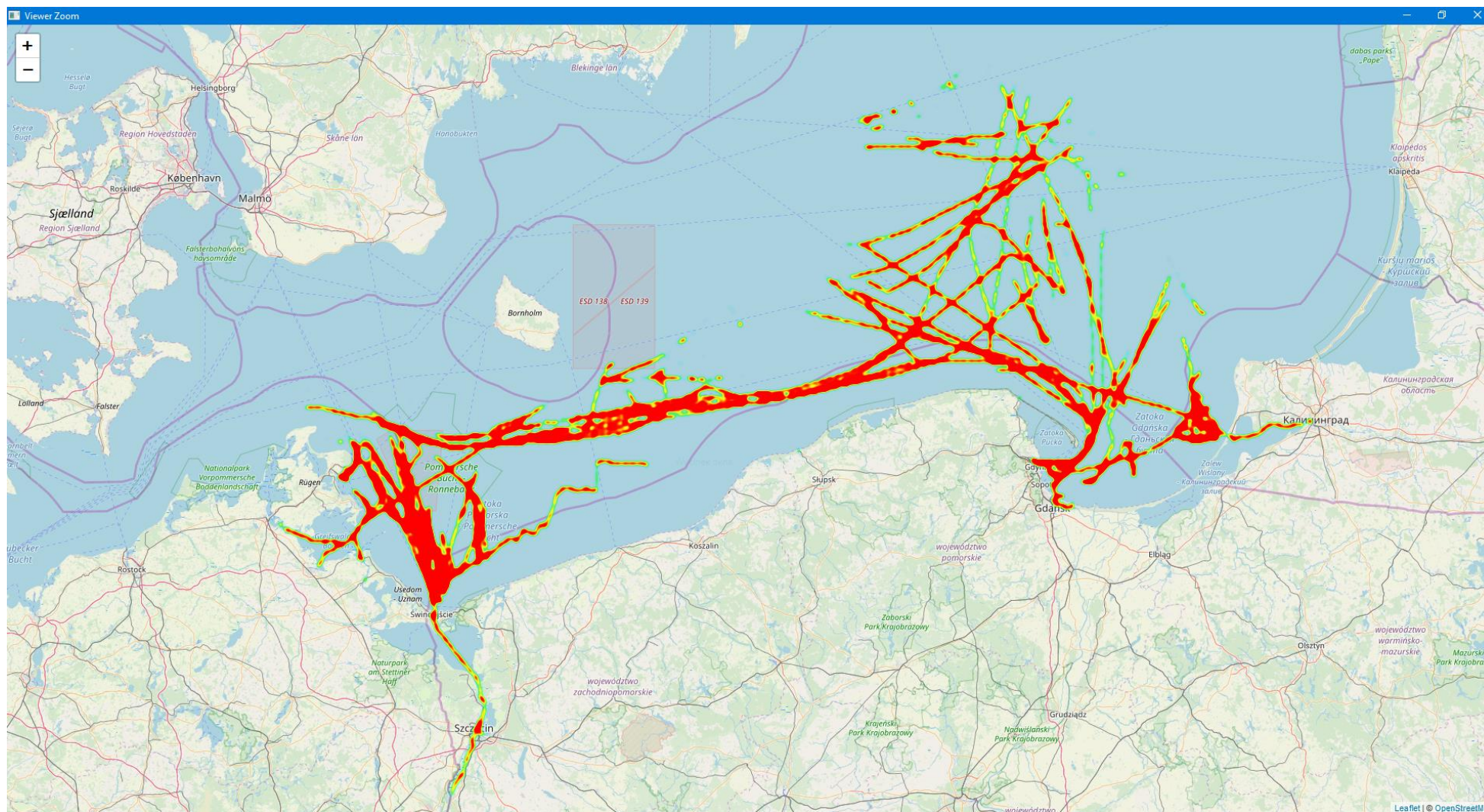
От веб-скраппинга и API до партнёрств по обмену данными с частным сектором — каждый этап включает проверку данных, контроль метаданных и мониторинг качества.



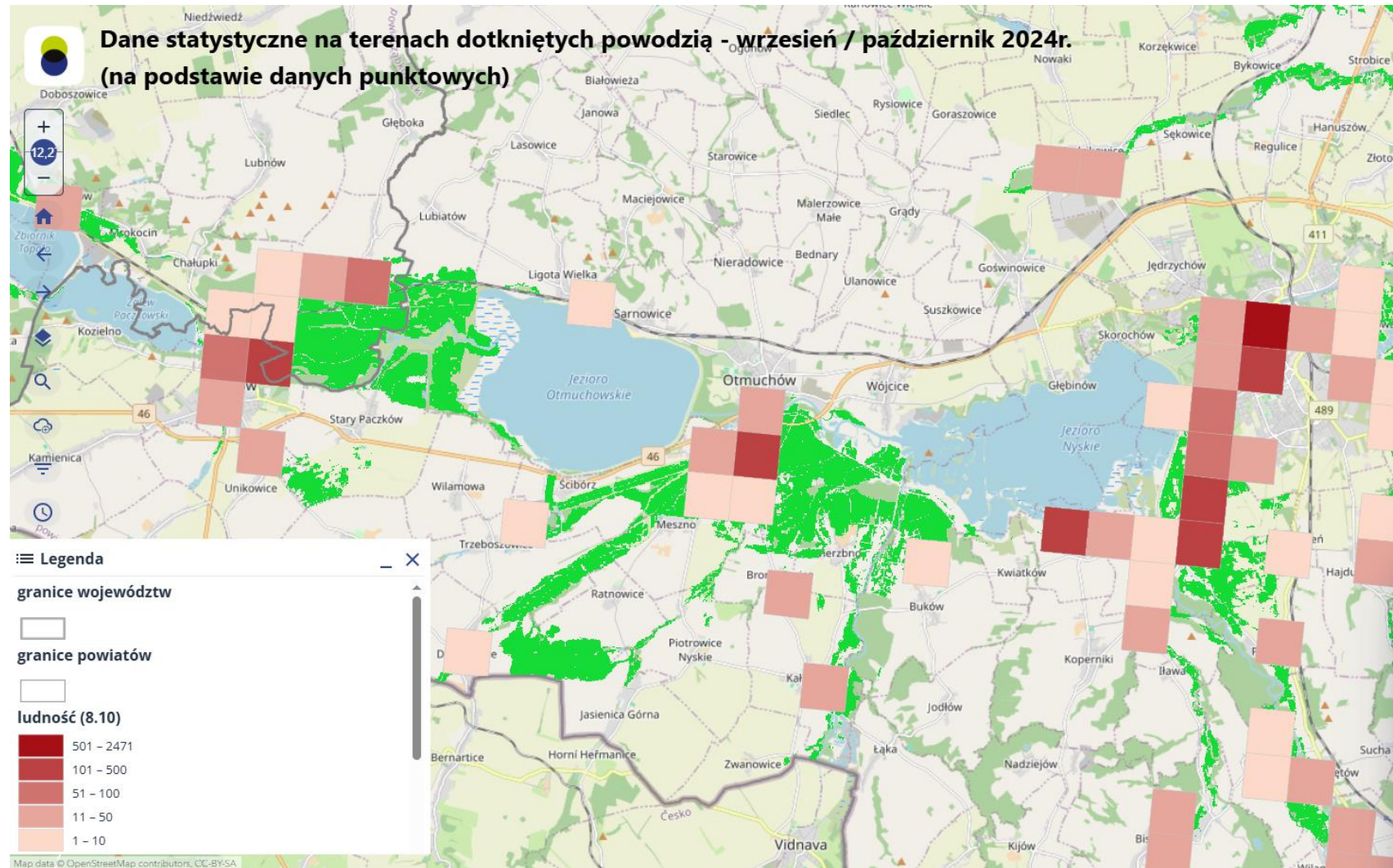
Разнообразие источников данных — разнообразие вызовов качества данных (статистика в области туризма)



Разнообразие источников данных — разнообразие вызовов качества данных (статистика в области судоходства)



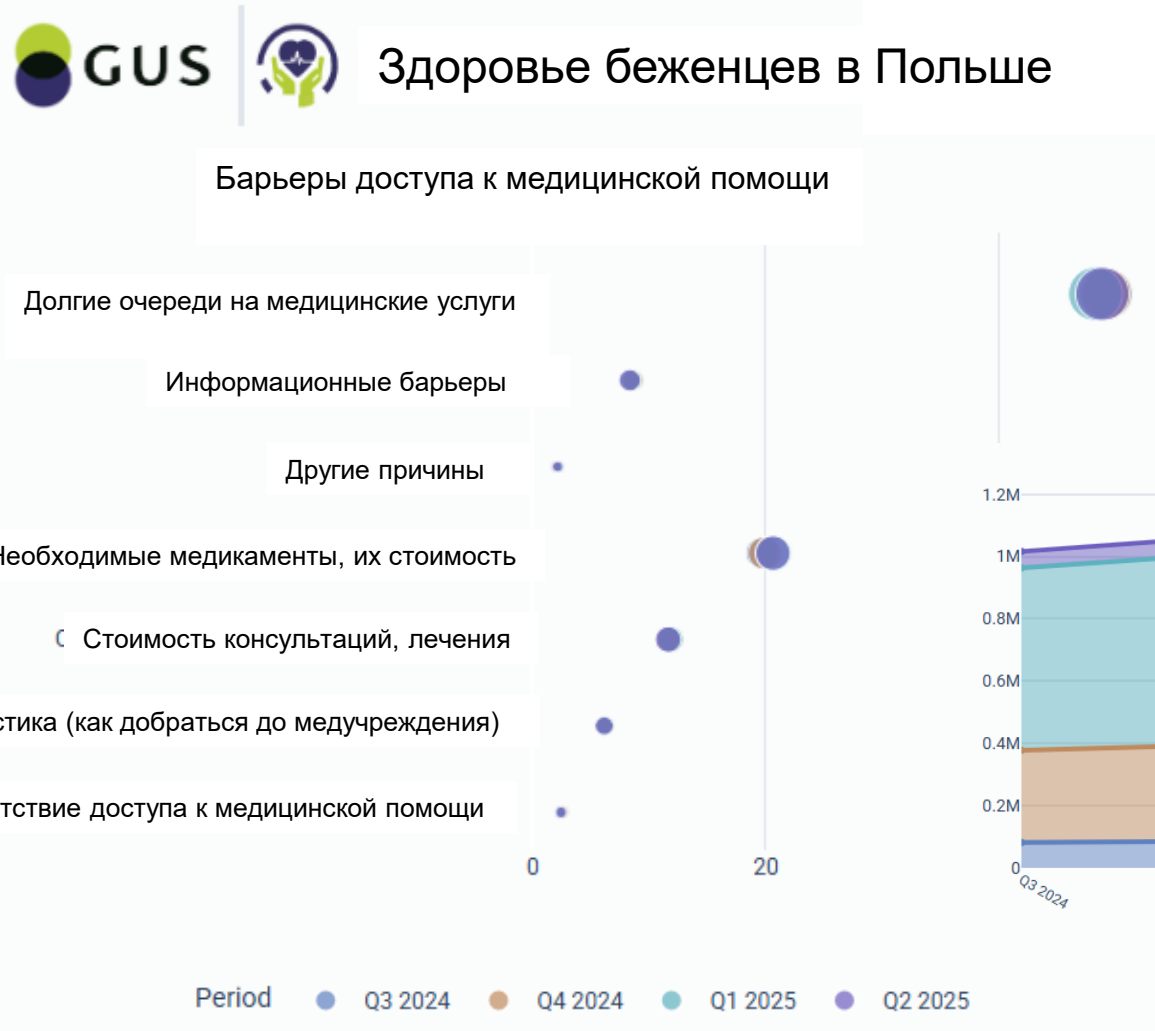
Разнообразие источников данных — разнообразие вызовов качества данных (районы, пострадавшие от наводнений)




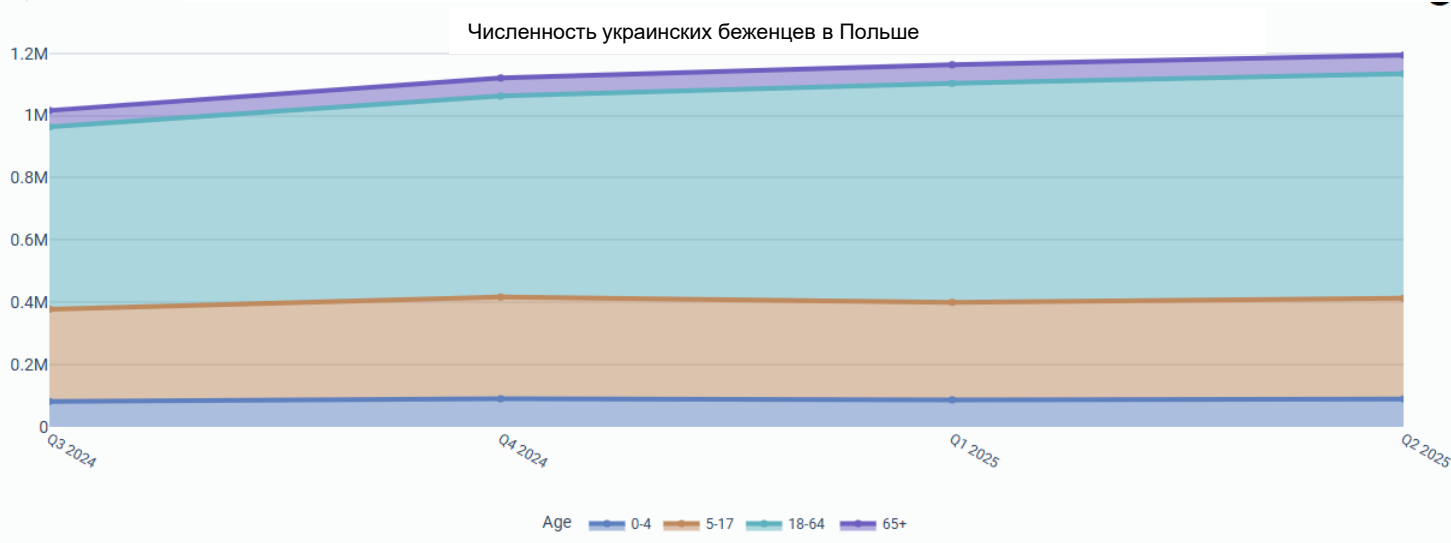
Данные спутниковой съёмки — распознавание сельскохозяйственных культур



Данные операторов мобильной связи



 **Очень хорошо!**



Большие данные и качество данных в официальной статистике

От объёма к ценности — качество как основа доверенных инноваций

Источники данных

- веб-данные
- спутниковые изображения
- IoT
- датчики

Техники

- веб-скраппинг
- машинное обучение
- автокодирование и классификация
- восполнение данных
- оценки

Примеры применения

- онлайн-вакансии
- характеристики предприятий
- статистика сельского хозяйства
- городские территории
- статистика туризма
- надёжная умная статистика
- качество данных
- covid-19



Большие данные: возможности, вызовы и риски для качества данных

ВЫГОДЫ

- Снижение затрат на исследования и нагрузки на респондентов
 - Сокращение сроков предоставления данных, вплоть до (почти) оперативной аналитики в реальном времени
 - Новая информация, новые компетенции в государственной статистике
-

ВЫЗОВЫ

- Сотрудничество с поставщиками данных из частного сектора
 - Законодательное регулирование отстаёт от технологического прогресса
 - Интеграция с существующими системами государственной статистики, сопоставимость данных
 - Обеспечение методологической обоснованности новых источников
 - Адаптация ИТ-инфраструктуры
-

РИСКИ

- Плата за доступ к данным
- Статистическая конфиденциальность, защита персональных данных
- Качество и стабильность данных, недостаточная полнота информации
- Отток специалистов в частный сектор
- Неопределённость качества и отсутствие прозрачности частных источников данных

Большие данные: возможности, вызовы и риски для качества данных

ВЫГОДЫ

- Снижение затрат на исследования и нагрузки на респондентов
- Сокращение сроков предоставления данных, вплоть до предоставления в реальном времени
- Новая информация, новые компетенции в государстве

ВЫЗОВЫ

- Сотрудничество с поставщиками данных из частного сектора
- Законодательное регулирование отстаёт от технологического прогресса
- Интеграция с существующими системами государственной статистики, сопоставимость данных
- Обеспечение методологической обоснованности новых источников
- Адаптация ИТ-инфраструктуры

РИСКИ

- Плата за доступ к данным
- Статистическая конфиденциальность, защита персональных данных
- Качество и стабильность данных, недостаточная полнота информации
- Отток специалистов в частный сектор
- Неопределённость качества и отсутствие прозрачности частных источников данных

ДАШБОРДЫ

Большие данные: возможности, вызовы и риски для качества данных

ВЫГОДЫ

- Снижение затрат на сбор и обработку данных, снижение нагрузки на респондентов
- Сокращение сроков получения информации вплоть до получения информации в реальном времени
- Новая информация, новые источники данных

Академия Data Science

ДАШБОРДЫ

ВЫЗОВЫ

- Сотрудничество с поставщиками данных из частного сектора
- Законодательное регулирование отстаёт от технологического прогресса
- Интеграция с существующими системами государственной статистики, сопоставимость данных
- Обеспечение методологической обоснованности новых источников
- Адаптация ИТ-инфраструктуры

РИСКИ

- Плата за доступ к данным
- Статистическая конфиденциальность, защита персональных данных
- Качество и стабильность данных, недостаточная полнота информации
- Отток специалистов в частный сектор
- Неопределённость качества и отсутствие прозрачности частных источников данных



Большие данные: возможности, вызовы и риски для качества данных

ВЫГОДЫ

- Снижение затрат на сбор и обработку данных, снижение нагрузки на респондентов
- Сокращение сроков получения информации вплоть до получения информации в реальном времени
- Новая информация, новые возможности

Академия Data Science

ДАШБОРДЫ

ВЫЗОВЫ

- Сотрудничество с поставщиками данных из частного сектора
- Законодательное регулирование отстаёт от технологического прогресса
- Интеграция с существующими системами государственной статистики
- Обеспечение методологической обоснованности новых источников данных
- Адаптация ИТ-инфраструктуры



THREATS

- Плата за доступ к данным
- Статистическая конфиденциальность, защита персональных данных
- Качество и стабильность данных, недостаточная полнота информации
- Отток специалистов в частный сектор
- Неопределённость качества и отсутствие прозрачности частных источников данных



Большие данные: возможности, вызовы и риски для качества данных

ВЫГОДЫ

- Снижение затрат на сбор и обработку данных, снижение нагрузки на респондентов
- Сокращение сроков получения информации вплоть до получения информации в реальном времени
- Новая информация, новые возможности

Академия Data Science

ДАШБОРДЫ

CHALLENGES

- Сотрудничество с поставщиками данных из частного сектора
- Законодательное регулирование отстаёт от технологий
- Интеграция с существующими системами
- Обеспечение методологической обоснованности данных
- Адаптация ИТ-инфраструктуры

WEB INTELLIGENCE HUB

EUROSTAT

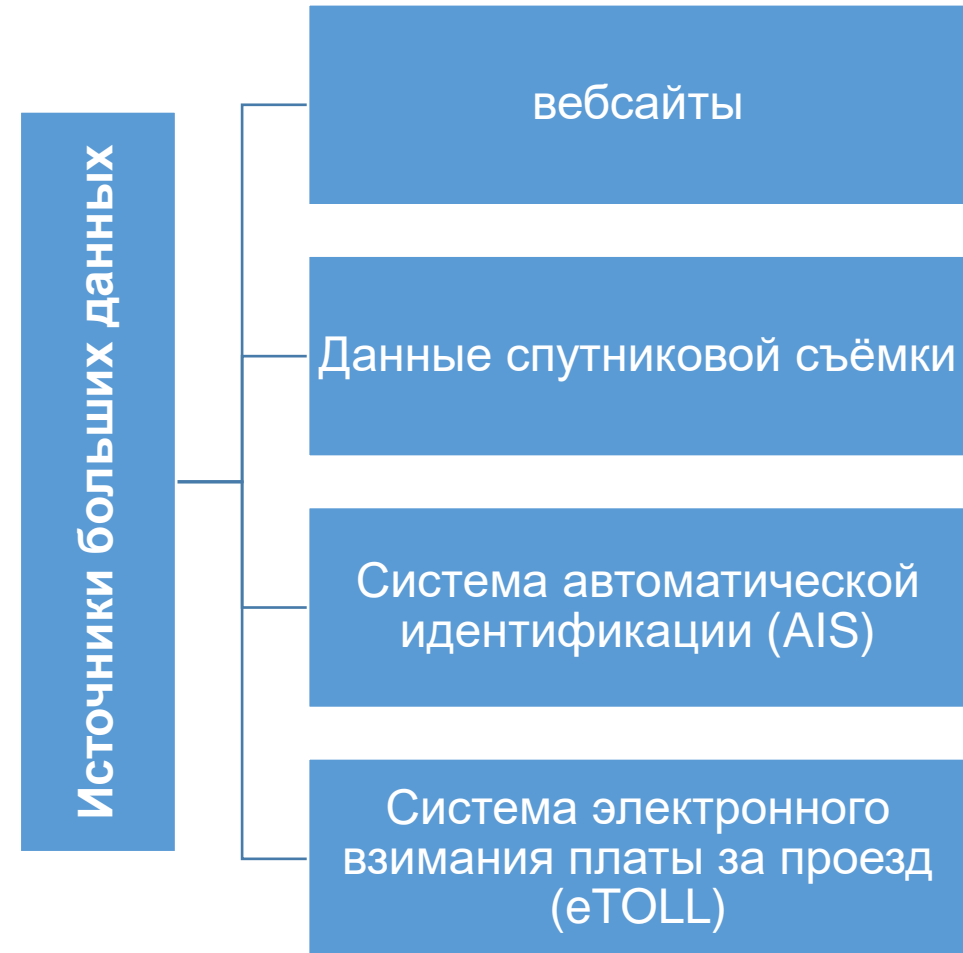
ФОРУМ ОБМЕНА ДАННЫМИ

ВЫЗОВЫ

- Плата за доступ к данным
- Статистическая конфиденциальность, защита персональных данных
- Качество и стабильность данных, недостаточная полнота информации
- Отток специалистов в частный сектор
- Неопределённость качества и отсутствие прозрачности частных источников данных



Источники данных



Ключевые измерения качества данных в цепочке создания данных

Точность и
надёжность

Своевременность и
согласованность

Прозрачность и
стандарты
метаданных

Этичное
использование и
обеспечение
конфиденциальности

Повторное
использование и
долгосрочная
ценность

Взгляд в будущее — качество данных в современном мире

- Данные, готовые к ИИ
- Методология оценки качества больших данных и моделей машинного обучения
- Непрерывное развитие ИТ - системы при выполнении текущих задач
- От сбора до коммуникации — качество встроено на каждом этапе цепочки создания ценности данных

Спасибо!

Сіздерге рақмет

Доминика Рогалиньска,

D.Rogalinska@stat.gov.pl

**Качественные данные
создают доверие.
Доверие создаёт
ценность.**