



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ



Институт статистических исследований и экономики знаний НИУ ВШЭ

# ТЕНДЕНЦИИ ИМПОРТОЗАМЕЩЕНИЯ В ИТ: РЕЗУЛЬТАТЫ АНАЛИЗА БОЛЬШИХ ДАННЫХ ГОСЗАКУПОК РФ



Илья Кузьминов  
Директор центра стратегической  
аналитики и больших данных  
ИСИЭЗ НИУ ВШЭ  
[ikuzminov@hse.ru](mailto:ikuzminov@hse.ru)

Онлайн-конференция "Импортозамещение в ИТ 2021"  
Москва, 25.03.2021



# СОДЕРЖАНИЕ

---

- О системе анализа текстовых и документных данных iFORA
- Источники данных о закупках в России и за рубежом
- Проведенные подготовительные работы
- Возможности анализа данных
- Анализ уровня импортозамещения продуктов по данным госзакупок

**> 400** млн  
документов

**+30** тыс.  
документов ежедневно

## Языки

Английский    +Латинские  
Русский        +Кириллические  
Китайский

- > 300 тыс. Отчеты о НИР
- > 250 млн. Научные публикации
- > 1 млн. Гранты
- > 100 тыс. Научные конференции
- > 100 млн. Патенты
- > 5 тыс. Образовательные программы
- > 2 млн. Вакансии
- > 2 млн. Нормативная правовая база
- > 1 млн. Документы международных организаций, консалтинговых компаний
- > 53 млн. Рыночная аналитика и профессиональные СМИ
- > 300 тыс. Научные проекты международных и национальных программ / фондов



### Российская апробация:

Научно-технический совет Военно-промышленной комиссии РФ  
Более 250 презентаций  
Более 30 докладов на конференциях

### Стратегическая аналитика

(стратегии, прогнозы, приоритеты, долгосрочные программы развития, программы инновационного развития, инвестиционные программы, технологические дорожные карты и др.)

### Международная апробация:

OECD (Париж), NISTEP (Токио)  
Global TechMining Conference (Атланта)  
Joanneum Research (Вена), University of Manchester и др.



В 2018 г. проект iFORA™ отмечена ОЭСР в качестве **успешной инициативы в области цифровизации науки** (OECD Science, Technology and Innovation Outlook 2018)



В 2020 г. iFORA™ упомянута в журнале **Nature** в качестве эффективного инструмента поддержки принятия решений (Nature, 2020, Vol. 583)



## Аналитический конструктор

Тренды	Оценки рынков	Прогнозы	Бенчмаркинг и оценки рисков	Анализ политики	Сети и центры компетенций	Компетенции и навыки	Проектное управление
Картирование трендов	Количественные оценки рынков	Драйверы, барьеры, эффекты	Анализ конкурентоспособности организаций, продуктов, технологий	Анализ документов политики	Выявление специализации организаций	Анализ образовательных программ	Оперативная конкурентная разведка для принятия решений
Анализ структурных изменений	Выявление возникающих технологий	Таймлайны событий будущего	Репутационный анализ на основе фактов	Выявление приоритетов	Анализ кооперации и аффилиаций	Выявление востребованных компетенций	Анализ закупок
Анализ жизненного цикла	Технологические и продуктовые портфели	Агрегация прогнозных оценок	Сентимент- и контент-анализ	Legal AI, анализ пробелов в нормативной базе	Анализ экспертного ландшафта	Оценка влияния технологий на профессии	Экспертиза проектов, проектных заявок
Поиск ответов на большие вызовы	Выявление зарождающихся рынков	Количественные консенсус-прогнозы	Оценки рисков	Сопоставление российской и международной повестки	Выявление центров и сетей компетенций (организаций и специалистов)	Прогнозирование спроса на компетенции	Технологические и управленческие дорожные карты
Анализ влияния технологий на сектора	Анализ региональных рынков	Прогнозирование развития технологий / продуктов / услуг	Индикаторы воздействия СМИ и рекламы			Выявление возникающих профессий	Формирование проектных команд
	Идентификация стартапов		Независимые рейтинги				

**Система iFORA основана на модульном подходе и позволяет комбинировать специализированные модули для конкретных задач**

**1** Выбор профильных источников информации из базы данных с возможностью ее пополнения

Полные тексты на разных языках

Разнообразные форматы данных

Отбор источников по единым объективным критериям

Постоянное пополнение базы данных

**2** Обработка больших данных, обучение предиктивных моделей, поиск неявных связей, выявление закономерностей

Уникальное сочетание подходов к текст-майнингу

Компьютерная лингвистика

Анализ больших данных и машинное обучение с элементами ИИ

Онтологическая инженерия

**3** Формирование аналитических материалов

Прозрачная, воспроизводимая методика анализа

Снижение рисков "человеческого фактора"

Высокая детализация аналитических результатов

Результатом являются витрины данных

**4** Экспертная валидация и формирование рекомендаций

Первичная интерпретация карт

Углубленный анализ детализирующей информации

Формирование выводов и рекомендаций



# Источники данных о закупках в России и за рубежом



## Россия



Единая информационная система в сфере закупок  
[www.zakupki.gov.ru](http://www.zakupki.gov.ru)



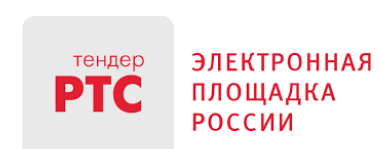
Система торгов Сбербанк-АСТ  
[www.sberbank-ast.ru](http://www.sberbank-ast.ru)



Единая электронная торговая площадка  
[www.roseltorg.ru](http://www.roseltorg.ru)



Общероссийская система электронной торговли  
[www.zakazrf.ru](http://www.zakazrf.ru)



Электронная площадка России  
[www.rts-tender.ru](http://www.rts-tender.ru)



Федеральная электронная площадка «НЭП – Фабрикант»  
[www.etp-ets.ru](http://www.etp-ets.ru)



Всероссийская универсальная площадка  
[www.lot-online.ru](http://www.lot-online.ru) и др.

## США



Federal Procurement Data System  
[www.fpds.gov](http://www.fpds.gov)

## КНР



中央政府采购网  
[www.zycg.gov.cn](http://www.zycg.gov.cn)



Неправительственная организация в сфере закупок  
[www.bidchance.com](http://www.bidchance.com)

## Канада



Public Works and Government Services Canada

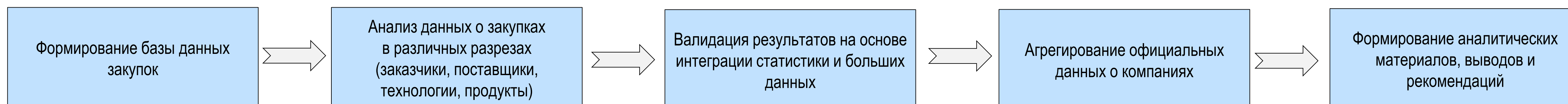
Public Works and Government Services Canada  
[www.buyandsell.gc.ca](http://www.buyandsell.gc.ca)

## Великобритания



Contracts Finder  
[www.contractsfinder.service.gov.uk](http://www.contractsfinder.service.gov.uk) и др.

## Анализ данных из разнородных источников



## СОЗДАННАЯ БАЗА ДАННЫХ О ЗАКУПКАХ

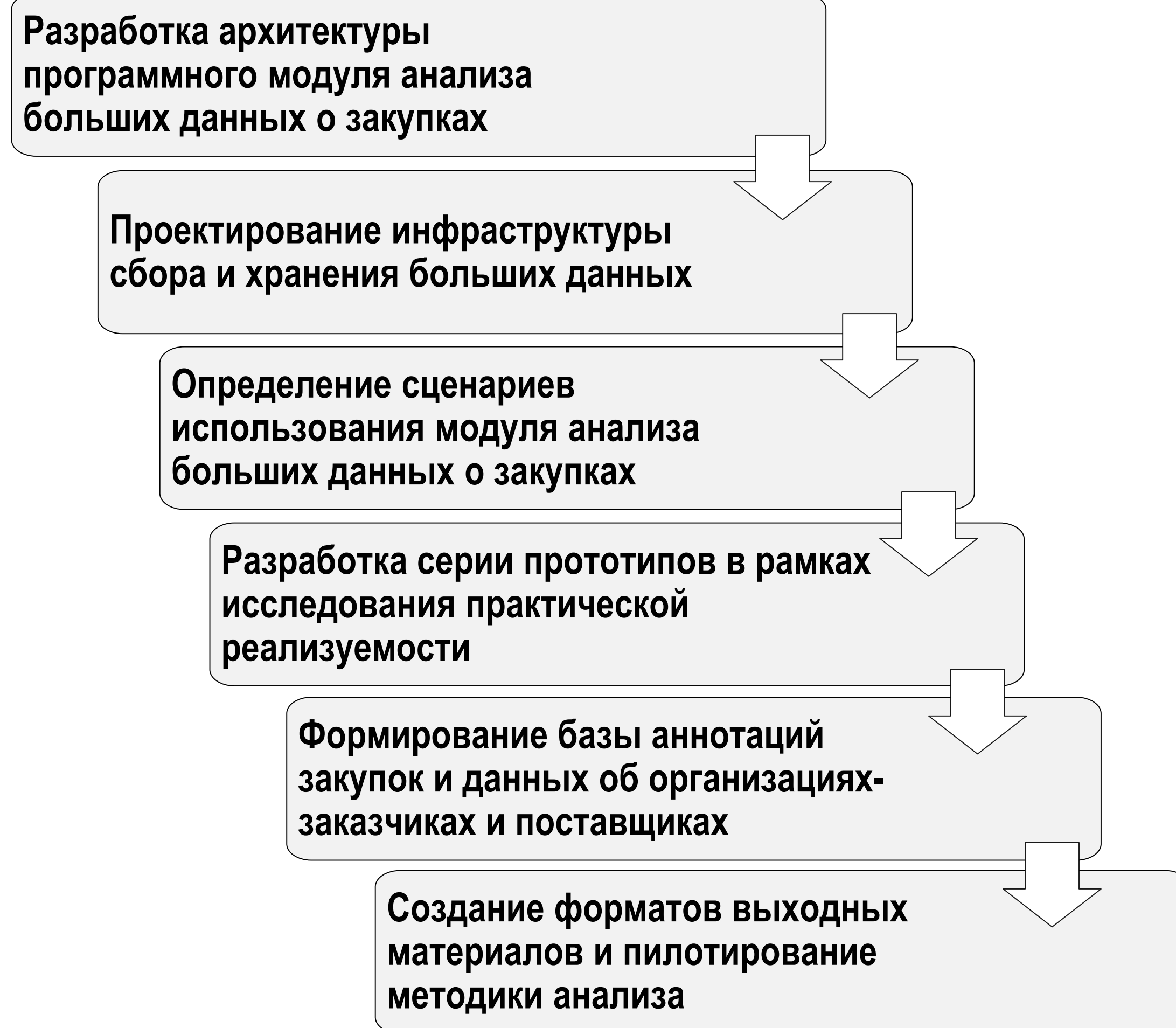


### Единая информационная система в сфере закупок

Источником для тестового анализа стали открытые данные, официально предоставляемые оператором Единой информационной системы в сфере закупок: [www.zakupki.gov.ru](http://www.zakupki.gov.ru)

Объектами тестового анализа стали компании, работающие по закону 223-ФЗ

- 6,2 млн аннотаций закупок за период 01.01.2015 - 09.03.2020 на общую сумму 101,97 трлн. руб.
- 39 тыс. заказчиков и 70 тыс. поставщиков
- ИТ-закупки: 2,3% от общего числа – 140,7 тыс. на общую сумму 1,38 трлн. руб. (1,4%)



интерфакс МЕЖДУНАРОДНАЯ ИНФОРМАЦИОННАЯ ГРУППА

### Информационный ресурс СПАРК



Информационный ресурс СПАРК, доступный в Едином архиве экономических и социологических данных НИУ ВШЭ, стал источником дополнительных необходимых сведений о анализируемых организациях-заказчиках и поставщиках

## Задачи углубленного интеллектуального анализа больших данных:

- проанализировать контракты в разрезе технологий и закупок ПО
- получить количественные данные для статистического анализа

## Выявленные препятствия:

- неконсистентность доступных перечней ПО и других продуктов на основе цифровых технологий
- неограниченное разнообразие вариантов наименований технологий и продуктов
- отсутствие форматно-логического контроля полей данных на стороне источника информации

Разработанный алгоритм анализа основан на **двухкомпонентном ансамбле моделей машинного обучения:**

### 1. Нейросетевая модель бинарной классификации коротких текстов

- Задача – определять вероятность отнесения тематик в названиях и аннотациях закупок к цифровым технологиям
- Модель обучена на более 3 тыс. примеров, достигнутый уровень точности (ROC AUC) – **99%**

### 2. Нейросетевая модель извлечения именованных сущностей

- Задача – выявлять названия продуктов в названиях и аннотациях закупок
- Модель обучена на более 2 тыс. примеров, достигнутый уровень качества: точность (Precision) – **84%**, полнота (Recall) – **87%**, агрегированный показатель качества (F-Score) – **85%**



# Интеллектуальный анализ и интеграция данных из ресурса СПАРК позволили расширить набор доступных параметров закупок



## Ключевые параметры закупок, доступные в данных:

- реквизитные данные контрактов, даты заключения
- реквизитные данные 39 тыс. заказчиков и 70 тыс. Исполнителей по контрактам
- цены и валюты контрактов
- объекты закупок



## На основе сбора дополнительных открытых данных и семантической обработки документов закупки также классифицированы в разрезе:

- цифровых технологий
- продуктовых групп, классов и наименований ПО
- производителей продуктов



## Для углубленного анализа данные из системы закупок обогащены сведениями о компаниях из информационного ресурса СПАРК, включая:

- вид деятельности и организационно-правовая форма
- финансовые показатели (активы, выручка, прибыль)
- регион регистрации, возраст компаний и др.

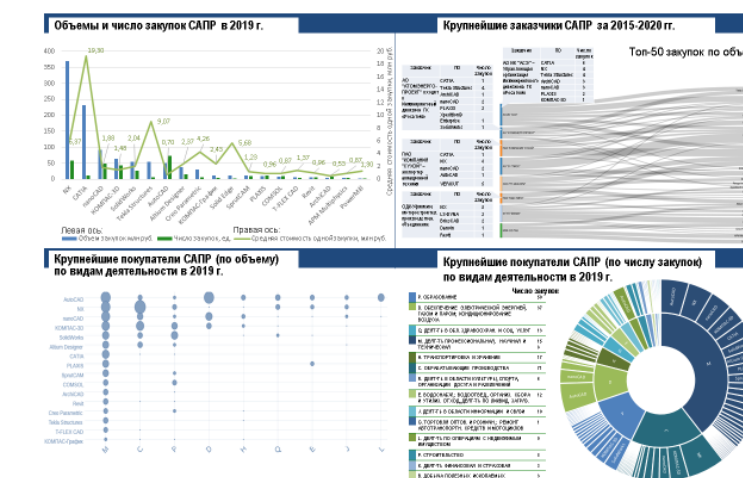
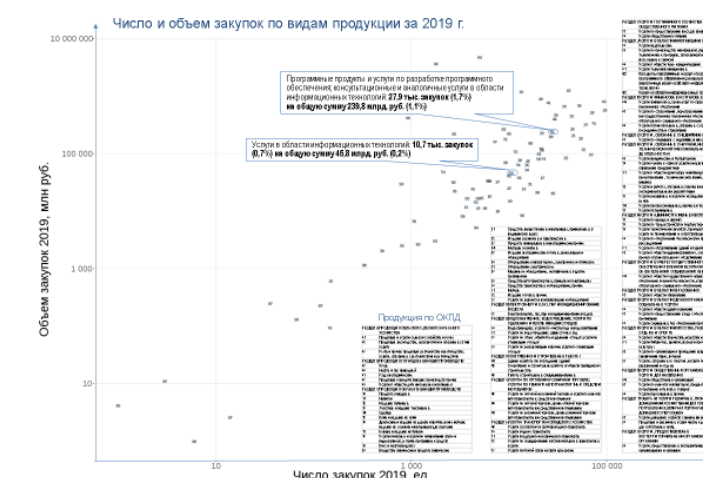


## Возможности анализа данных:

- по видам продукции (по ОКПД и технологиям)
- по компаниям-заказчикам и поставщикам (по ОКВЭД и др. параметрам)
- по регионам и странам

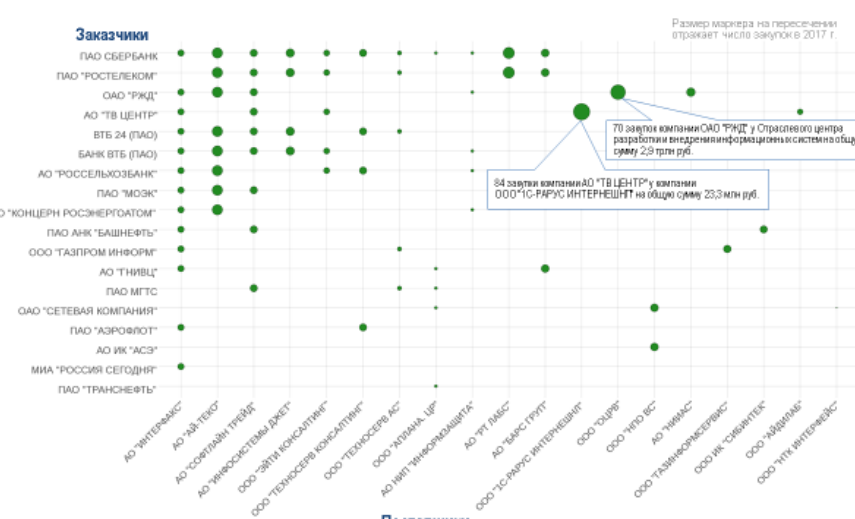
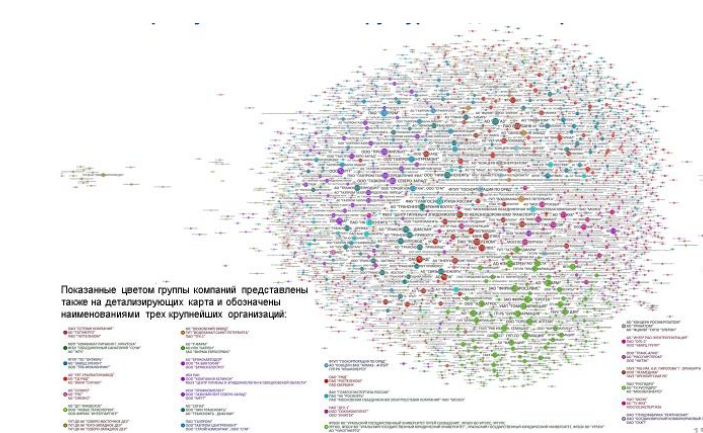
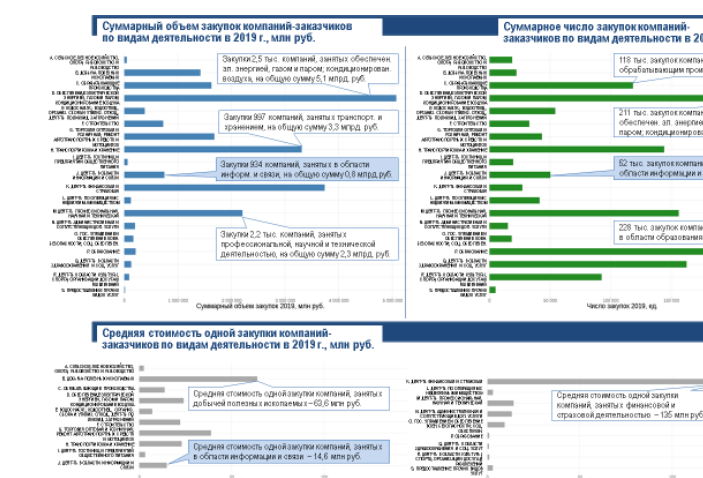
## По видам продукции:

- Целесообразность анализа закупок того или иного типа продуктов, определение доли конкретного продукта в общем объеме закупок госкомпаний (например, ИТ-закупок)
- Комплексный анализ закупок какого-либо продукта (например, анализ САПР как вида ПО)
- Анализ уровня импортозамещения какого-либо продукта



## По компаниям-заказчикам и поставщикам:

- Анализ закупок в зависимости от характеристик компаний-заказчиков, в т.ч. в разрезе видов деятельности
- Анализ закупок в зависимости от характеристик компаний-поставщиков, в т.ч. в разрезе видов деятельности
- Поиск связей между заказчиками и поставщиками может быть, его детализация до уровня отдельного продукта
- Гибкое картирование связей между заказчиками и поставщиками в целях представления комплексной картины сложившейся структуры отдельных рынков



## По регионам и странам:

- Сопоставление параметров закупок между регионами в целях дополнительного среза ключевых тенденций

Средний объем закупок компаний по регионам\* за 2019 г., млн руб.



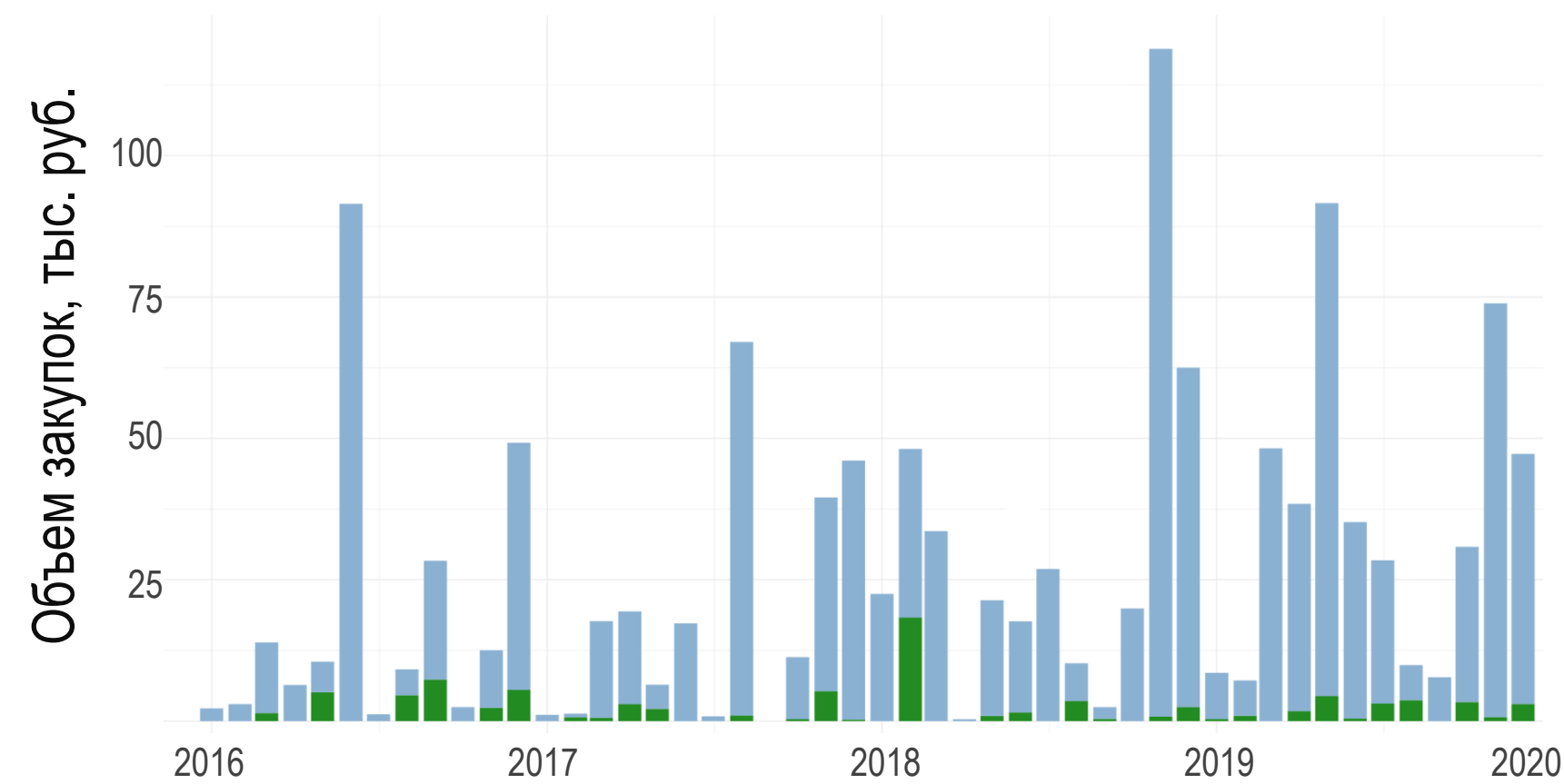
\* Данные по Архангельской и Тюменской областям не учитываются ввиду отсутствия информации о точном объеме. Показатели округлены в пользу соответствующей области.

Топ-10 регионов по среднему объему закупок компаний (отношению суммарного объема закупок к числу компаний-заказчиков) в 2019 г.

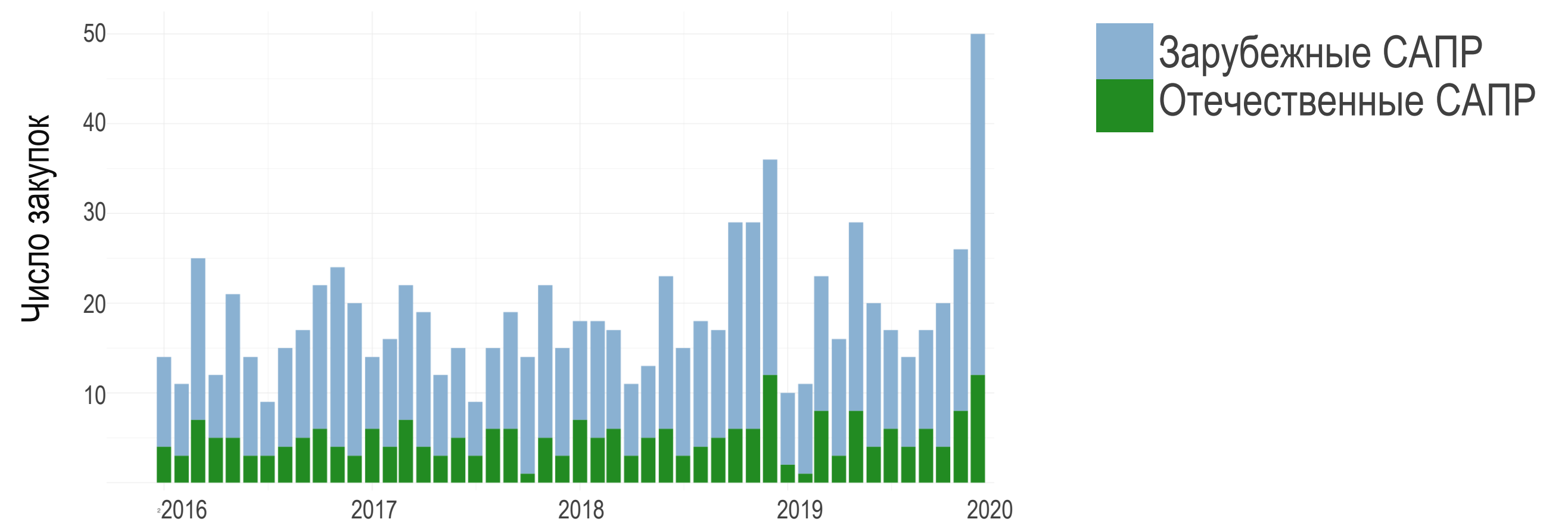
Регион	Средний объем закупок, млн руб.	Среднее число закупок	Число компаний-заказчиков
г. Москва	5 496,4	190,3	1 912
Московская область	2 427,9	62,6	1 428
г. Санкт-Петербург	2 158,1	136,2	608
Татарстан	1 821,9	86,9	140
Республика Татарстан	1 673,1	112,9	224
Республика Саха (Якутия)	730,5	79,0	223
Калужская область	724,0	62,4	178
Рязанская область	651,7	65,1	122
Свердловская область	640,7	37,0	1 571
Приморский край	593,4	50,0	264

# Анализ уровня импортозамещения ИТ-продуктов на примере систем автоматизированного проектирования (САПР) в 2016-2020 гг.

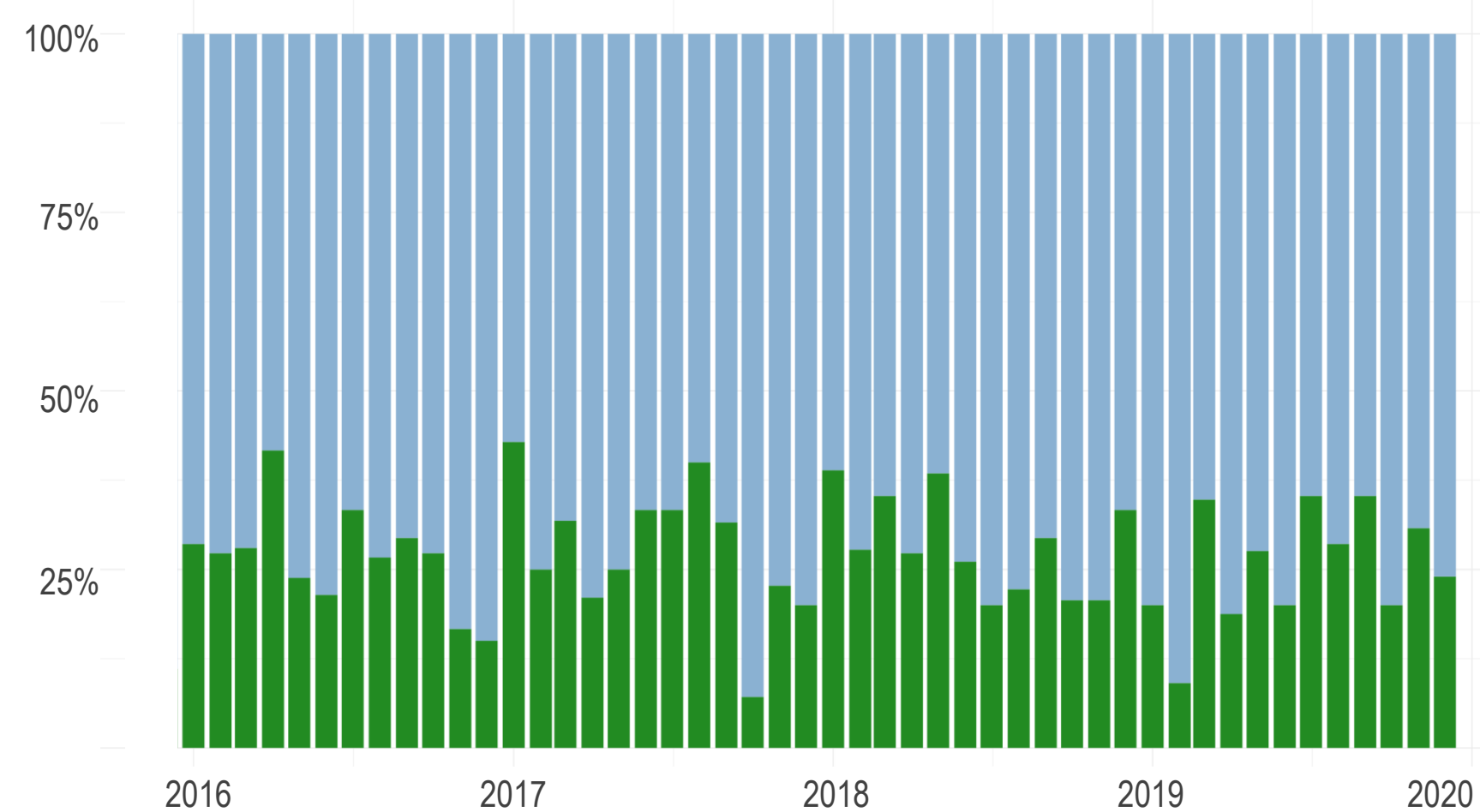
Динамика объема закупок зарубежных и отечественных САПР программ



Динамика числа закупок (в абсолютном выражении) зарубежных и отечественных САПР программ



Динамика числа закупок (в долевого выражении) зарубежных и отечественных САПР программ



Компании-производители САПР

ПО	Компания-производитель	Страна производства	Доля закупок (по числу) в 2019 г., %
AutoCAD	Autodesk	США	20,7
NX	Siemens PLM Software	США	16,4
nanoCAD	АО "НАНОСОФТ"	Россия	<b>13,9</b>
КОМПАС-3D	ООО "АСКОН-Системы проектирования"	Россия	<b>12,2</b>
SolidWorks	Dassault Systèmes	Франция	7,6
Altium Designer	Altium	США	4,5
CATIA	Dassault Systèmes	Франция	3,4
PLAXIS	Bentley Systems	США	3,1
SprutCAM	ООО "СПРУТ-ТЕХНОЛОГИЯ"	Россия	<b>2,5</b>
COMSOL	COMSOL	Швеция	2,5
ArchiCAD	Graphisoft	Венгрия	2,3
Creo Parametric	PTC	США	2,0
Revit	Autodesk	США	2,0
Tekla Structures	Tekla	Финляндия	1,7
КОМПАС-График	ООО "АСКОН-Системы проектирования"	Россия	<b>1,4</b>
T-FLEX CAD	ЗАО "ТОП СИСТЕМЫ"	Россия	<b>1,4</b>
APM Multiphysics	ООО Научно-технический центр "АГПМ"	Россия	<b>1,1</b>
Solid Edge	Siemens PLM Software	США	0,6
PowerMill	Autodesk	США	0,6



# Анализ уровня импортозамещения ИТ-продуктов на примере систем автоматизированного проектирования (САПР) в 2016-2020 гг.



Говорить о перспективах эффективного импортозамещения САПР пока рано

Совокупный стоимостной объем отечественных и зарубежных закупок САПР нестабилен

Совокупное число отечественных и зарубежных закупок САПР в среднем стабильно, отмечается рост в последний год

Доля зарубежных закупок САПР в структуре общего числа закупок САПР стабильно выше отечественных

Производители САПР представлены шестью странами

Стоимостной объем зарубежных закупок САПР стабильно выше отечественных

Число зарубежных закупок САПР стабильно выше отечественных

Доля отечественных закупок САПР в структуре общего числа закупок САПР в среднем стабильна и составляет от 10% до 40%

Топ-2 продукта САПР по доле закупок в структуре общего числа закупок САПР представлены США

Стоимостной объем отечественных закупок САПР стабильно незначителен

Число отечественных закупок САПР стабильно невысоко

Лидирующий САПР-продукт, представленный РФ (nanoCAD), входит в топ-3 рейтинга



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
УНИВЕРСИТЕТ

Спасибо за внимание!

[www.issek.hse.ru](http://www.issek.hse.ru)

Телефон: +7 495 772 95 90 доб. 12022

E-mail: [ikuzminov@hse.ru](mailto:ikuzminov@hse.ru)