

УДК 004:001.89

DOI 10.17513/snt.40362

ВОСПРОИЗВОДИМАЯ МЕТОДОЛОГИЯ НАУКИ О ДАННЫХ

^{1,2}Горбунов В.И., ¹Салимов Т.А.¹ФГАОУ ВО «Национальный исследовательский институт ИТМО»,
Санкт-Петербург, e-mail: gorbunov.v93@gmail.com;²ФГБОУ ВО «Санкт-Петербургский государственный университет»,
Санкт-Петербург, e-mail: gorbunov.v93@gmail.com

Современное состояние научной деятельности характеризуется стремительным ростом объемов цифровых данных, усложнением исследовательских процессов и увеличением междисциплинарных взаимодействий, что обуславливает необходимость переосмысления традиционной методологии проведения исследований. Целью настоящего исследования является разработка воспроизводимой методологии науки о данных, в которой научная деятельность представляется как совокупность формализованных процессов, взаимодействующих с центральным объектом – Информацией. Методология разработана на основе анализа актуальной научной литературы, философских оснований научного метода, а также опыта междисциплинарных проектов, ориентированных на воспроизводимые вычисления и системную работу с данными. В рамках работы структурированы семь воспроизводимых процессов: сбор наблюдений, сбор вопросов, сбор публикаций, исследование, создание концептуальных фреймворков, публикация и создание дизайна. Каждый из процессов определен через фазу анализа и синтеза и направлен на трансформацию элементов Информации, включающей определения, наблюдения, вопросы, концептуальные фреймворки, прогнозы и дизайн процессов. Методология позволяет отделить процессы производства знания от самого знания, что способствует повышению прозрачности, воспроизводимости и верифицируемости научной деятельности. Предложенная структура обеспечивает формализацию научной деятельности, ее управляемость и адаптивность к условиям цифровой науки, сохраняя преемственность классического научного метода.

Ключевые слова: наука о данных, воспроизводимость, концептуальный фреймворк, исследовательская методология, научный метод

REPRODUCIBLE DATA SCIENCE METHODOLOGY

^{1,2}Gorbunov V.I., ¹Salimov T.A.¹ITMO University, Saint Petersburg, e-mail: gorbunov.v93@gmail.com;²Saint Petersburg State University, Saint Petersburg, e-mail: gorbunov.v93@gmail.com

The current state of scientific activity is characterized by a rapid increase in the volume of digital data, the growing complexity of research processes, and the expansion of interdisciplinary interactions, which necessitates a reconsideration of traditional research methodology. The aim of this study is to develop a reproducible data science methodology in which scientific activity is represented as a set of formalized processes interacting with a central object – Information. The methodology was developed based on an analysis of relevant scientific literature, philosophical foundations of the scientific method, and the experience of interdisciplinary projects focused on reproducible computation and systematic data handling. The study structures seven reproducible processes: observation collection, question collection, publication collection, investigation, conceptual framework creation, publication, and design development. Each process is defined through phases of analysis and synthesis and is aimed at transforming elements of Information, which includes definitions, observations, questions, conceptual frameworks, predictions, and process designs. The methodology enables a separation between the processes of knowledge production and the knowledge itself, thereby enhancing the transparency, reproducibility, and verifiability of scientific activity. The proposed structure provides for the formalization, manageability, and adaptability of scientific work in the context of digital science, while preserving the continuity of the classical scientific method.

Keywords: scientific method, data science, reproducibility, conceptual framework, research methodology

Введение

Современная наука развивается в условиях экспоненциального роста объемов цифровых данных, что оказывает значительное влияние на организацию и методологию научной деятельности. Усложнение исследовательских процессов, а также расширение междисциплинарных взаимодействий требует пересмотра классических подходов к производству, интерпретации и проверке научного знания.

Одной из ключевых тенденций последних лет становится интеграция подходов науки о данных в структуру научных исследований [1]. Эта область, изначально возникшая на пересечении информатики, статистики и прикладных дисциплин, формирует новые требования к организации научной работы, включая обработку больших данных, автоматизацию анализа, повторяемость вычислений и стандартизацию понятийного аппарата. В этих условиях возникает потребность в формализованной

методологической рамке, которая могла бы объединить философию классического научного метода с инструментами, развивающимися в науке о данных [2].

Также актуален вопрос о четком разграничении между процессами, создающими знание, и самим знанием как объектом. Это позволяет повысить системность и управляемость научной деятельности, особенно в междисциплинарной среде, где необходимо согласование терминологии, процедур и методов между различными научными школами и подходами.

Целью исследования является разработка воспроизводимой методологии науки о данных, представленной как система формализованных процессов, взаимодействующих с центральным объектом – Информацией, и ориентированной на применение в междисциплинарной научной среде.

Материалы и методы исследования

Исследование имело теоретико-методологический характер и проводилось в период с 2022 по 2025 г. в контексте междисциплинарной профессиональной деятельности, связанной с анализом больших данных и применением методов машинного обучения.

Методологическая база формировалась на основе систематического обзора научной литературы, посвященной трансформации научного метода в условиях цифровизации науки. В качестве источников рассматривались рецензируемые научные статьи, методические материалы и публикации, описывающие инженерную практику работы с данными.

Среди ключевых теоретических оснований – концепция центрированности на данных (data centrism), трактующая научную деятельность как процесс системного накопления и преобразования информации, получаемой из наблюдений [3]. Также учитывались подходы, направленные на интеграцию машинного обучения в классическую структуру научного метода, при котором гипотезы и концепции уточняются в ходе анализа больших массивов данных [4]. Дополнительно рассматривались философские критерии научности, в частности тестируемость и фальсифицируемость как неотъемлемые условия научной объяснительной модели [5].

Для анализа и синтеза информации применялись методы тематической категоризации, создания онтологий, построения ментальных карт и формализация процессов в табличной и схематичной формах. Эти методики позволяли структурировать подходы, выделять общие принципы и фор-

мировать модель научной деятельности, устойчивую к междисциплинарным различиям. Каждый элемент и процесс был определен на основе анализа как классических научных практик, так и современных данных о поведении исследовательских коллективов, систем открытой науки и платформ для совместного анализа данных [6].

Результаты исследования и их обсуждение

Результатом проведенной работы является пересмотр и формализация методологии научной деятельности с опорой на философские и практические подходы. В рамках предложенной методологии объектом исследования науки о данных являются методы извлечения ценности из оцифрованных данных, отражающих изучаемое явление.

Методология позволяет рассматривать научное исследование как циклический процесс, в котором данные, понятия и объяснительные конструкции проходят через формальные и эмпирические стадии обработки. Ниже представлены компоненты этой методологии.

Объект *Информация* представляет собой ядро методологии – совокупность научных сущностей, формализующих знание, получаемое в процессе исследования. Все семь процессов методологии взаимодействуют с Информацией: наполняют, преобразуют и используют ее. В состав Информации входят следующие элементы: определения, наблюдения, вопросы, концептуальные фреймворки, прогнозы и дизайны процессов.

Рассмотрим каждый элемент Информации по порядку, обозначив терминологию, затем перейдем к рассмотрению места элемента Информации в общем процессе.

1. *Определения* – фундаментальная основа, на которой строятся все остальные сущности. Определения задают язык, структуру и логику всей исследовательской деятельности. Каждое наблюдение, вопрос, прогноз или элемент дизайна формируется и интерпретируется исключительно в рамках заданных определений. Это делает определения не просто вспомогательным, а системообразующим компонентом всей методологии.

Обсуждение, документирование и обмен знаниями в ходе исследования преимущественно осуществляются на естественном языке. Поскольку наука о данных является междисциплинарной областью, ее терминология отличается от используемых понятий в исследуемых предметных областях. В связи с этим важным аспектом организации научной деятельности является ведение и поддержание актуального глос-

сария, в котором фиксируются определения используемых терминов и методов. Четкость и полнота определений способствуют снижению количества проблем в коммуникациях как внутри исследовательской группы, так и во взаимодействии с внешними специалистами.

Определения делятся на два типа:

– Формальные определения – выражаются в строго определенной логической или математической форме [7]:

▪ Аксиомы – базовые утверждения, принимаемые без доказательств.

▪ Правила вывода – правил, с помощью которых (и только с помощью их) из аксиом и ранее выведенных теорем можно образовывать новые теоремы.

▪ Теоремы – выводы, строго вытекающие из аксиом по правилам вывода.

▪ Меры и метрики – характеристики для оценки свойств объектов.

▪ Измерения – числовые характеристики по результатам замера свойств объектов для фиксированного набора параметров измерения.

– Естественные определения – формулируются на естественном языке и отражают понятийную базу исследуемой области:

▪ Термины предметной области (например, сопоставление терминов из науки о данных с терминами предметной области). Научные определения терминов, которые могут отличаться от их обыденного употребления в естественном языке, что требует фиксации их значений для исключения неоднозначности.

▪ Методы и процедуры (например, «качественное интервью», «АБ тестирование»).

▪ Инструменты (например, «электронный микроскоп», «опросная анкета»).

Требования к определениям:

– Непротиворечивость – логическая и терминологическая согласованность.

– Полнота – достаточность информации для однозначного понимания и применения.

– Трансляция – возможность сопоставления внешних определений с внутренними.

2. *Наблюдения* – зафиксированные проявления эмпирического мира. Они интерпретируются только в рамках существующих определений.

Наблюдения классифицируются:

– по форме:

▪ качественные (описательные характеристики);

▪ количественные (числовые измерения);

– по источнику:

▪ первичные (полученные в ходе текущего исследования);

▪ вторичные (взяты из внешних публикаций);

– по стадии обработки:

▪ сырые (необработанные, полученные в результате сбора наблюдений);

▪ преобразованные (результаты предварительного анализа, нормализации и агрегации) [8].

Сырые наблюдения служат основой для проверки прогнозов, а преобразованные используются при построении моделей и будущих исследований.

3. *Вопросы* – формализованные направления научного поиска. Они задают границы и цели исследования, инициируют построение концептуальных фреймворков, а также требуют точного понятийного аппарата. Вопросы могут быть как прикладными, так и теоретическими (например, «Как изменяется X при Y?» или «Какие факторы определяют Z?»).

Исследовательские вопросы служат связующим звеном между наблюдаемым миром и будущими объяснениями. Они задают направления поиска, определяют релевантность информации и формируют основу для построения концептуальных фреймворков.

Вопросы могут быть ориентированы на уточнение терминов, проверку закономерностей, разработку методов измерения и т.д. Четкая формулировка вопросов позволяет задать направления поиска в рамках дальнейших процессов [9].

4. *Концептуальные фреймворки* – объяснительные структуры, объединяющие:

– законы – описывают наблюдаемые зависимости без объяснения причин;

– гипотезы – предполагаемые объяснения, подлежащие эмпирической проверке.

– теории – обобщения, интегрирующие несколько гипотез. Теория отличается от гипотезы степенью обоснования и масштабностью охвата.

Для научной значимости важно, чтобы концепт был не только проверяем, но и фальсифицируем. Принцип тестируемости требует, чтобы существовала возможность эмпирической проверки концепции, а принцип фальсифицируемости – чтобы концепция могла быть потенциально опровергнута.

Эти идеи восходят к философу Карлу Попперу, который утверждал, что научная концепция считается научной лишь в том случае, если она допускает опровержение. Концепция, которая не может быть опровергнута никакими данными, не поддается научной проверке и остается вне научного метода. Поэтому концепция должна формулироваться таким образом, чтобы ее можно было подвергнуть эмпирическому испытанию и потенциальному опровержению.

5. *Прогнозы* – логические следствия из концептуальных фреймворков, выраженные в терминах, зафиксированных в определениях, и предсказывающие наблюдения:

- формируются на основе существующих теорий и гипотез;
- служат инструментом эмпирической проверки теоретических конструкций через сравнение с сырыми наблюдениями [10];
- являются инструментом верификации концептуальных конструкций и в случае расхождения требуют их ревизии.

6. *Дизайн процессов* – формализованное описание будущей исследовательской деятельности [11]:

- может касаться любого процесса методологии: сбора наблюдений, построения модели, публикации и др.;
- включает цели, методы, метрики оценки, критерии успешности, ограничения и ресурсы;
- опирается на определения и другие сущности Информации;
- обеспечивает воспроизводимость, согласование действий исследователей и возможность предварительного анализа рисков.

Для формирования, использования и передачи Информации заданы семь процессов, которые можно разделить на три группы: поставляющие информацию, работающие с ней и управляющие исследовательской деятельностью. Рассмотрим подробнее каждый процесс.

1. *Сбор наблюдений* представляет собой процесс научной деятельности, направленный на фиксацию феноменов эмпирического мира.

Этот процесс включает два этапа:

- Анализ – определяется, какие частные признаки феномена будут зафиксированы, поскольку невозможно запечатлеть феномен в полном объеме. Это обусловлено как ограничениями органов чувств и инструментальных средств регистрации, так и тем, что любое наблюдение связано с выбором исследователя, который опирается на существующую систему знаний и целей исследования. Полное отображение феномена потребовало бы фиксации бесконечного множества его характеристик, что невозможно в рамках конечной системы представления знаний.

- Синтез – выбранные признаки объединяются с существующими определениями, что придает им субъективную интерпретацию, обусловленную структурой имеющихся знаний.

Наблюдение может осуществляться как в пассивной форме, так и с активным вмешательством, при котором исследователь изменяет параметры среды с целью на-

блюдения реакций системы. Такой подход приближается к эксперименту, но не включает немедленного вывода обобщений – интерпретация происходит на последующих этапах.

При фиксации наблюдений исследователь должен стремиться к объективности, избегая субъективных интерпретаций. Для повышения надежности наблюдений необходимо проводить повторный множественный сбор наблюдений разными людьми, а также учитывать вероятность случайных совпадений, что требует увеличения размера выборки.

Таким образом, формируется репрезентация объекта эмпирического мира в формальной системе, в рамках которой будет осуществляться дальнейшая работа.

2. *Сбор вопросов* происходит параллельно с фиксацией наблюдений и формирует пространство исследовательских вопросов. Это отдельный процесс в рамках методологии, выполняющий функцию формализации исследовательских интересов и постановки задач, на которые предполагается найти объяснение или решение. В отличие от сбора наблюдений, основанного на фиксации эмпирических данных, сбор вопросов работает как с эмпирическим миром, обозначая интересующие свойства феноменов, так и с уже накопленной информацией, выявляя области неполноты знания, которые требуют дальнейшего объяснения, уточнения или проверки.

Процесс сбора вопросов состоит из двух фаз:

- Анализ – исследователь выбирает интересующие аспекты феномена либо просматривает накопленную информацию с целью выявления элементов, вызывающих интерес или порождающих неопределенность. Вопрос никогда не формулируется в абсолютной пустоте, он опирается на уже имеющиеся понятия, определения, параметры и границы исследования.

- Синтез – после того, как выделены ключевые элементы, наступает переход от фокусировки на проблемной зоне к явной формулировке вопроса.

3. *Сбор публикаций* представляет собой систематизированный процесс интеграции внешнего знания в структуру текущего исследования. Научные публикации выступают важнейшим источником вторичной информации: они содержат результаты предыдущих наблюдений, сформулированные концепции, методы, а также оформленные термины и метрики. Включение внешнего материала требует сопоставления понятий и интерпретации в контексте существующей Информации [12].

Сбор публикаций – это специализированный процесс в рамках методологии, отличающийся от сбора наблюдений и вопросов своей природой: он не ориентирован на фиксацию эмпирических феноменов, а направлен на интеграцию существующих научных знаний, зафиксированных в работах других исследователей. Если сбор наблюдений связан с внешним эмпирическим миром, а сбор вопросов – с постановкой новых исследовательских задач, то сбор публикаций взаимодействует с уже оформленным знанием, прошедшим процедуры научной верификации, рецензирования и концептуализации. Это знание представлено в виде статей, монографий, тезисов, баз данных и других источников и может содержать определения, модели, концепции, методы, метрики, а также обоснованные выводы и примеры интерпретации.

Процесс сбора публикаций включает две фазы:

– Анализ – на этом этапе происходит целенаправленный отбор публикаций, релевантных исследуемой теме. Критерии анализа могут включать: тематическое соответствие, научную значимость и уровень цитируемости, применяемые методы, точность и формальность определений, новизну и обоснованность выдвигаемых концепций. Анализ требует предварительного понимания структуры Информации, уже сформированной в рамках исследования: отбор осуществляется с учетом существующих определений, концептуальных фреймворков и поставленных вопросов. Это позволяет исключить нерелевантные источники и сконцентрироваться на тех, что могут быть встроены в текущую методологическую схему.

– Синтез – после отбора публикаций начинается фаза интерпретации и адаптации полученного материала. Задачи синтеза включают:

- извлечение ключевых определений, методов, гипотез и метрик;
- сопоставление терминологии внешних источников с внутренними определениями;
- оценку согласованности новых концептов с существующими;
- выявление противоречий, уточнение или обоснованное расширение понятийного аппарата;
- формирование новых сущностей в объекте Информация (например, включение новых метрик в раздел «определения» или новых гипотез в концептуальный фреймворк).

Синтез может также сопровождаться аннотированием, категоризацией или переводом выбранного материала в структурированный формат (например, глоссарии, библиографические базы, онтологии).

Процесс сбора публикаций играет ключевую роль в обеспечении преемственности научного знания и повышении обоснованности новых концептуальных построений. В рамках методологии он, наряду со сбором наблюдений и вопросов, предоставляет информацию, необходимую для последующих исследований.

4. *Исследование* представляет собой процесс применения и развития формальных методов для анализа существующей Информации. Его цель – интерпретация наблюдений, проверка прогнозов, оценка гипотез и построение логически непротиворечивых, воспроизводимых научных выводов в виде формальных и естественных определений. Исследование не занимается созданием новых концептуальных фреймворков или прогнозов – это отдельная задача, выполняемая в процессе создания концептуальных фреймворков. Вместо этого исследование опирается на уже сформированные концепции и прогнозы, проверяя их внутреннюю логическую согласованность, применимость и соответствие наблюдаемым данным [13, 14]. Исследование состоит из двух фаз – анализа и синтеза.

– Анализ – исследователь изучает существующие определения, концептуальные фреймворки и прогнозы, выявляя объекты, требующие формального рассмотрения. При этом могут быть поставлены задачи логической верификации концептуальных фреймворков, проверки соответствия между прогнозом и сырыми наблюдениями, уточнения метрик или оценки применимости конкретного метода. Анализ также может выявить ограничения используемых инструментов или необходимость в разработке новых формальных подходов.

– Синтез – исследователь осуществляет непосредственное применение или развитие формальных методов. В рамках этого этапа исследователь может разрабатывать математические модели, строить алгоритмы анализа, проводить статистическую обработку, формулировать логические выводы или выполнять симуляции. Важным аспектом синтеза является не только вычисление или оценка параметров (измерение), но и изложение полученных результатов в терминах исходных определений и понятийной структуры Информации, а также сохранения артефактов, необходимых для воспроизведения измерения.

Таким образом, в процессе исследования осуществляется как прикладная аналитика, направленная на верификацию существующих утверждений, так и теоретическая работа, связанная с расширением инструментального аппарата самой методологии.

Исследование играет центральную роль в формализации научной деятельности. Оно обеспечивает воспроизводимость, точность и логическую непротиворечивость полученных результатов, позволяет проводить проверку научных гипотез и прогнозов в формализованной среде и тем самым поддерживает связь между теоретическими построениями и эмпирическими наблюдениями.

5. *Создание концептуальных фреймворков и прогнозов* – это процесс, направленный на формулирование новых объяснительных конструкций, позволяющих интерпретировать наблюдаемые явления и связывать их с существующими определениями и вопросами [15], а также создавать на их основе прогнозы для проверки согласованности с наблюдениями. Это единственный процесс в методологии, в рамках которого формируются новые гипотезы, теории, законы и прогнозы. Его результатом является структурированное объяснение, которое может быть использовано для дальнейшего анализа и проверки. Процесс включает две фазы: анализ и синтез.

– Анализ – исследователь изучает имеющуюся Информацию: наблюдения, определения, вопросы, уже имеющиеся концептуальные фреймворки. Основная задача этого этапа – выявить логические пробелы, неформализованные участки знания или паттерны в наблюдениях, которые не объясняются существующими концептуальными фреймворками. Также на этом этапе проводится сопоставление различных понятий, оценка согласованности между наблюдениями и текущими концепциями, выявление повторяющихся закономерностей или аномалий, которые требуют объяснения. Анализ может опираться как на внутреннюю структуру Информации, так и на знания, полученные в ходе сбора публикаций (после их сопоставления с существующими определениями и терминологией).

– Синтез – формулируются новые концептуальные фреймворки. Это может включать построение объяснительных моделей, выведение гипотез, разработку теоретических структур и введение новых определений на естественном языке. В рамках этой фазы также создаются прогнозы – конкретные, проверяемые следствия из предложенных фреймворков, выраженные в виде предсказаний будущих наблюдений. Прогнозы представляют собой неотъемлемую часть концептуального фреймворка, поскольку именно через сравнение прогнозов с сырыми наблюдениями осуществляется проверка его научной состоятельности.

Создание концептуального фреймворка – это творческая и аналитическая деятельность, направленная на переход от эмпирического накопления данных к построению объяснительных моделей при помощи дедуктивных и индуктивных подходов к формированию научного знания. Этот процесс важен, поскольку в формальной части методологии работают с уже заданными структурами, тогда как формулирование самих концептуальных рамок требует выхода за пределы существующего формального аппарата. Законы, входящие в структуру фреймворка, описывают устойчивые эмпирические зависимости, но не дают объяснений их причин. Гипотезы, в отличие от законов, предполагают объяснение, но требуют их проверки. Теории представляют собой систематизированное и многократно проверенное объяснение, объединяющее несколько гипотез и законов.

Таким образом, процесс создания концептуальных фреймворков и прогнозов тесно связан с другими процессами: использует определения, вопросы и наблюдения, а также предоставляет материал для последующего формального анализа в процессе исследования и для планирования дальнейшей деятельности в рамках создания дизайна. Создание фреймворков не просто оформляет знание – оно задает структуру, в рамках которой знание может быть проверено, уточнено и воспроизведено.

6. *Публикация* – это процесс структуризации накопленного научного знания в форме, пригодной для передачи другим исследователям и включения в дальнейшую научную деятельность. Это не просто фиксация результатов – публикация замыкает цикл процессов воспроизводимого исследования, обеспечивая внешнюю проверку, распространение и возможность повторного анализа. В методологии публикация рассматривается как процесс, основанный на двух фазах: анализе и синтезе.

– Анализ – исследователь переосмысливает накопленную Информацию, отбирая те элементы, которые подлежат передаче вовне. Этот этап включает определение того, какие наблюдения, определения, прогнозы, фреймворки и результаты исследования являются завершенными, значимыми и логически связанными между собой. Также на этом этапе оценивается, в каком виде информация может быть воспроизведена другими: рассматривается необходимость включения метрик, ссылок на данные, открытого кода и схем процессов. Анализ направлен на вычленение логически цельных блоков знаний, которые могут быть проверены и использованы независимо.

– Синтез – осуществляется структурированное оформление выбранной информации. Это может включать создание научной статьи, отчета, препринта, модели, онтологии или базы данных. Важно не только изложить суть полученного знания, но и представить его в такой форме, которая обеспечивает воспроизводимость, верифицируемость и совместимость с другими исследованиями. Здесь фиксируются определения, методы, параметры, прогнозы, исходные данные и используемые концептуальные фреймворки. Публикация может сопровождаться аннотацией, описанием модели, приложением кода или ссылкой на наборы данных, что делает возможной повторную реализацию представленного материала.

Публикация тесно связана с другими процессами. Она использует Информацию, сформированную в ходе всего исследовательского цикла: определения, наблюдения, концептуальные фреймворки, прогнозы, результаты исследования. Она может ссылаться на публикации, собранные в процессе сбора внешних источников, и, в свою очередь, становится частью Информации в других исследованиях, через механизм сбора публикаций. Кроме того, публикация может быть заранее спланирована и структурирована в процессе создания дизайна – например, как ожидаемый результат определенного этапа работы.

Таким образом, публикация – это ключевой процесс фиксации и передачи научного знания. Она завершает цикл процессов воспроизводимого исследования и одновременно создает входные данные для следующих исследований: как источник новых определений, наблюдений, вопросов и концепций. Именно через публикации наука сохраняет преемственность, устойчивость и возможность коллективного наращивания знания.

7. *Создание дизайна* – это процесс формализации и планирования будущих действий в рамках научной деятельности. В отличие от других процессов, направленных на получение, анализ или представление знания, создание дизайна выполняет управленческую функцию: оно задает структуру, последовательность и условия выполнения других процессов. Дизайн позволяет заранее определить, какие шаги будут предприняты, с какими целями и по каким правилам, обеспечивая тем самым прозрачность, согласованность и воспроизводимость научной работы. Как и другие процессы методологии, создание дизайна включает две фазы: анализ и синтез.

– Анализ – исследователь оценивает текущую структуру Информации – уже сформулированные определения, наблюдения, вопросы, концептуальные фреймворки, прогнозы, а также результаты проведенных исследований. Цель анализа – понять, какие элементы требуют развития, какие вопросы остаются открытыми, какие методы можно применить и какие ресурсы необходимы для дальнейшей работы. Также на этом этапе учитываются ограничения: технические, методологические, временные или ресурсные. Анализ направлен на осознание как потенциальных направлений исследования, так и условий их реализации.

– Синтез – формулируется структурированный план научной деятельности. Это может быть дизайн сбора наблюдений (с указанием что, как и когда фиксировать), план сбора публикаций (темы исследования, открытые вопросы, пробелы в знаниях, источники публикаций), план сбора вопросов (чек-листы, опросные листы, реестр заинтересованных лиц), план исследования (методы, входные параметры и метрики), план создания концептуальных фреймворков (открытые вопросы, список участников, повестка встреч) и план создания публикации (структура, ожидаемые результаты, формат).

Дизайн описывает цели, предполагаемые действия, используемые определения и методы, критерии оценки успеха, а также процедуры контроля качества для обеспечения воспроизведения процесса и полученных результатов. Он может быть выражен в виде протокола, схемы, документа спецификаций, технического задания или даже кода. Создание дизайна обеспечивает возможность согласования действий между членами исследовательской группы, повышает управляемость проекта, снижает вероятность ошибок и позволяет сравнивать запланированное с фактически выполненным. Кроме того, наличие формализованного дизайна существенно упрощает воспроизводимость исследования: другие исследователи могут не только повторить действия, но и критически оценить выбор методов, метрик и процедур.

Таким образом, создание дизайна – это инструмент управления научной деятельностью. Оно превращает исследование из последовательности интуитивных решений в формализованный процесс, поддающийся планированию, контролю, анализу и повторному выполнению. Именно благодаря дизайну возможна коллективная и воспроизводимая организация научной работы в условиях растущей сложности и междисциплинарности современного научного знания.

Заключение

Предложенная методология представляет собой переосмысление научного метода в условиях цифровизации, междисциплинарности и стремительного роста объемов данных. Разделение процессов и объекта исследования (Информации) позволяет формализовать этапы научной деятельности, сделать их воспроизводимыми, гибкими и совместимыми с инструментами науки о данных.

Каждый из процессов выполняет уникальную функцию в рамках методологии: сбор наблюдений и вопросов обеспечивает насыщенность информационного пространства; исследование и создание концептуальных фреймворков – аналитическую и объяснительную деятельность; публикация и сбор публикаций – интеграцию с научным сообществом; создание дизайна – управление научной деятельностью.

Методология акцентирует внимание на важности четкого определения понятий, строгой формализации, интерпретируемости и проверяемости научных результатов. Она сохраняет преемственность классического научного метода, одновременно расширяя его за счет вычислительных и информационных практик науки о данных. Таким образом, предложенная методология может служить основой для построения воспроизводимых, масштабируемых и теоретически обоснованных научных программ будущего.

Дальнейшее развитие методологии может включать разработку инструментов автоматизации для документирования каждого из процессов, создание онтологий для формализованного описания информационных сущностей, а также интеграцию с платформами открытой науки и воспроизводимых вычислений. Также перспективным является применение методологии в образовательной среде – для формирования у исследователей системного понимания научной деятельности и навыков проведения воспроизводимых исследований.

Список литературы

1. Ezer D., Whitaker K. Data science for the scientific life cycle // *Nature Methods*. 2019. Vol. 16, Is. 7. P. 541–545. URL: DOI: 10.7554/eLife.43979 (дата обращения: 01.02.2025).
2. Carballo-Meilan A., McDonald L., Pragot W., Starowski L.M., Saleemi A.N., Waheed Afzal W. Development of a Data-Driven Scientific Methodology // *Chemometrics and Intelligent Laboratory Systems*. 2022. Vol. 227. Article ID: 104589. DOI: 10.1016/j.chemolab.2022.104555.
3. Stadelmann T., Klamt T., Merkt P.H. Data Centricism and the Core of Data Science // *Philosophy & Technology*. 2022. Vol. 8, Is. 2.; URL: <https://publikationen.bibliothek.kit.edu/1000143637> (дата обращения: 01.02.2025). DOI: 10.5445/IR/1000143637.
4. Peters D.P.C., Havstad K.M., Cushing J., Tweedie C., Fuentes O., Villanueva-Rosales N. Harnessing the power of big data: infusing the scientific method with machine learning to transform ecology // *Ecosphere*. 2014. Vol. 5, Is. 6. Article ID: art. 67. DOI: 10.1890/ES13-00359.1.
5. Elliott K.C., Cheruvil K.S., Montgomery G.M., Sorano P.A. Conceptions of Good Science in Our Data-Rich World // *BioScience*. 2016. Vol. 66, Is. 10. P. 880–889. DOI: 10.1093/biosci/biw115.
6. Succi S., Coveney P.V. Big data: the end of the scientific method? // *Philosophical Transactions of the Royal Society A*. 2019. Vol. 377, Is. 2142. Article ID: 20180145. DOI: 10.1098/rsta.2018.0145.
7. Deshpande D., Chhugani K., Ramesh T., Pellegrini M., Shiffman S., Abedalthagafi M.S., Alqahtani S., Ye J., Liu X.S., Leek J.T., Brazma A., Ophoff R.A., Rao G., Butte A.J., Moore J.H., Katritch V., Mangul S. The evolution of computational research in a data-centric world // *Cell*. 2024. Vol. 187, Is. 17. P. 4449–4457. DOI: 10.1016/j.cell.2024.07.045.
8. Maskey M. Rethinking AI for Science: An Evolution From Data-Driven to Data-Centric Framework // *Perspectives of Earth and Space Scientists*. 2023. Vol. 4, Is. 1. Article ID: e2023CN000222. DOI: 10.1029/2023CN000222.
9. Ramachandran R., Bugbee K., Murphy K. From Open Data to Open Science // *Earth and Space Science*. 2021. Vol. 8, Is. 5. Article ID: e2020EA001562. DOI: 10.1029/2020EA001562.
10. O'Brien E., Mick J. In the Academy, Data Science Is Lonely: Barriers to Adopting Data Science Methods for Scientific Research // *Harvard Data Science Review*. 2024. Vol. 6, Is. 2. DOI: 10.1162/99608f92.7ca04767.
11. Liu J., Carlson J., Pasek J., Puchala B., Rao A., Jagadish H.V. Promoting and Enabling Reproducible Data Science Through a Reproducibility Challenge // *Harvard Data Science Review*. 2022. Vol. 4, Is. 3. DOI: 10.1162/99608f92.9624ea51.
12. Stodden V. The data science life cycle: a disciplined approach to advancing data science as a science // *Commun. ACM*. 2020. Vol. 63, Is. 7. P. 58–66. DOI: 10.1145/3360646.
13. Karpatne A., Atluri G., Faghmous J.H., Steinbach M., Banerjee A., Ganguly A., Shekhar S., Samatova N., Kumar V. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data // *IEEE Transactions on Knowledge and Data Engineering*. 2017. Vol. 29, Is. 10. P. 2318–2331. DOI: 10.1109/TKDE.2017.2720168.
14. Borrego A., Dessi D., Ayala D., Hernández I., Osborne F., Reforgiato Recupero D., Buscaldi D., Ruiz D., Motta E. Research hypothesis generation over scientific knowledge graphs // *Knowledge-Based Systems*. 2025. Vol. 315. Article ID: 113280. DOI: 10.1016/j.knosys.2025.113280.
15. Ciardi T.G., Nihar A., Chawla R., Akanbi O., Tripathi P.K., Wu Y., Chaudhary V., French R.H. Materials data science using CRADLE: A distributed, data-centric approach // *MRS Communications*. 2024. Vol. 14, Is. 4. P. 601–611. DOI: 10.1557/s43579-024-00616-6.