



Pflichtenheft

Kurzübersicht

Projektname / Gruppenname:	ResearchGrid EU (Gruppe 4)	
Projektnummer:	004-RG-EU-2025	
Projektauftraggeber:	Enzo Hilzinger und Janett Betz	
Gruppenmitglieder:	Carlo Rinderer (Matr.-Nr.: 1902925) David Simon (Matr.-Nr.: 1893552) Niklas Seither (Matr.-Nr.: 4253802) Yanick Bedel (Matr.-Nr.: 8424886)	
Projektdauer:	Projektstartdatum:	09. Mai 2025
	Projektenddatum:	11. Juli 2025

Historie des Dokuments

Version	Datum	Autor	Bemerkung
1.0	21.05.2025	Gesamte Gruppe 4	Ersterstellung

Inhaltsverzeichnis

1. Einleitung	1
2. Detaillierte Zieldefinition	2
3. Datenquellen.....	5
4. Analysekonzepte der Hypothesen	7
4.1. Hypothese 1	7
4.2. Hypothese 2	8
4.3. Hypothese 3	8
4.4. Hypothese 4	9
4.5. Hypothese 5	9
4.6. Hypothese 6	10
4.7. Hypothese 7	10
5. Systemarchitektur und Projektstruktur	10
6. Storytelling-Ansatz	12
7. Technologien und Tools.....	13
8. Qualitätssicherung	14
9. Risiken und Gegenmaßnahmen	15
10. Bemerkungen	16
11. Finale Freigabe	16
12. Quellenverzeichnis.....	17

1. Einleitung

Dieses Pflichtenheft beschreibt die konkrete Umsetzung des im Lastenheft definierten Projekts zur Analyse von Einflussfaktoren auf den Erfolg wissenschaftlicher Forschung in Europa. Aufbauend auf den dort beschriebenen Zielsetzungen und Anforderungen wird im Folgenden dargelegt, wie die einzelnen funktionalen und nicht-funktionalen Anforderungen im Projektverlauf realisiert werden sollen.

Ziel der Umsetzung ist es, die im Lastenheft formulierten Thesen durch datenbasierte Analysen zu überprüfen. Dabei wird untersucht, ob und in welchem Ausmaß ein Zusammenhang zwischen dem Erfolg von Forschungsaktivitäten und den verschiedenen im Lastenheft definierten Einflussfaktoren besteht. Anhand von quantitativen Metriken wie Publikationsanzahl und Zitierungen, sollen diese Ergebnisse festgehalten werden. Im Fokus stehen insbesondere drei potenzielle Erfolgsfaktoren. Dazu gehört die Höhe der bereitgestellten finanziellen Mittel, die Beteiligung von Industriepartnern sowie die institutionelle Kooperation zwischen wissenschaftlichen Einrichtungen. Die Thesen zu diesen Faktoren werden systematisch operationalisiert und mit Hilfe öffentlich zugänglicher Forschungsdaten geprüft.

Die Analysen stützen sich dabei auf strukturierte Datenquellen wie OpenAlex und Eurostat. Die Datenauswertung erfolgt unter Verwendung etablierter Methoden der Datenanalyse und Visualisierung, die jeweils auf das Anwendungsgebiet zugeschnitten werden.

Die gewonnenen Ergebnisse werden so aufbereitet, dass sie auch für fachfremde Zielgruppen verständlich und nachvollziehbar sind. Hierzu wird eine sogenannte Data Story entwickelt, die die Analyseergebnisse in eine stringente, narrativ aufgebaute Struktur überführt und mit passenden Visualisierungen ergänzt. Zusätzlich erfolgt eine multimediale Aufbereitung in Form eines Informationsvideos, welches die Kernaussagen des Projekts kompakt und anschaulich kommuniziert. Beide Formate sollen sicherstellen, dass die Erkenntnisse des Projekts nicht nur wissenschaftlich fundiert, sondern auch kommunikativ wirkungsvoll vermittelt werden können.

Dieses Pflichtenheft stellt somit die verbindliche Grundlage für die technische und inhaltliche Umsetzung des Projekts dar und dient der klaren Ableitung von Analyse-, Visualisierungs- und Dokumentationsschritten im weiteren Projektverlauf.

2. Detaillierte Zieldefinition

Für die Analyse der Erfolgsfaktoren der europäischen Forschung wurden durch das Lastenheft zwei zentrale Untersuchungspunkte vorgegeben. Die sind durch die Handlungsmöglichkeiten der EU definiert und sollen demnach ausführlich untersucht werden. Um aussagekräftige Analysen im Rahmen dieser Vorgaben durchführen zu können, werden hierfür aus den definierten Bereichen präzise Hypothesen formuliert.

Im Rahmen der Untersuchung bezüglich der Auswirkungen der verfügbaren Mengen an monetäre Ressourcen der europäischen Forschung werden demnach folgende Fragestellungen untersucht:

- 1. Höhere Ausgaben in Forschung und Entwicklung führen zu größerem Forschungserfolg*
- 2. In den westlichen Mitgliedstaaten der europäischen Union zeigen die eingesetzten Forschungsgelder einen höheren Wirkungsgrad in Bezug auf den Forschungserfolg*
- 3. Elite-Institutionen erzeugen Ergebnisse, welche einen größeren Forschungserfolg verzeichnen*

Hypothese 1 richtet sich an die Grundlegende Annahme, dass mehr Ressourcen zu mehr beziehungsweise besseren Ergebnissen in der Forschung führen. Die Ergebnisse der Hypothese liefern als einen sehr allgemeinen Einblick in die Effektivität dieses Erfolgsfaktors und definiert damit eine gewisse Basis, welche durch die weiteren Hypothesen dieses Erfolgsfaktors näher untersucht werden soll. Die zweite Hypothese versucht sich im Anschluss an eine absolute Untersuchung des Forschungserfolges an einer Untersuchung der relativen Zusammenhänge zwischen Ressourcen und Forschungserfolg. Hierbei wird untersucht, wie effektiv Institutionen der einzelnen Mitgliedstaaten der EU ihre Mittel einsetzen, um so eine länderspezifische Idee davon zu bekommen, in welchen Ländern zukünftige Förderungen unter Umständen effektiver eingesetzt werden könnten, um den allgemeinen Forschungserfolg der EU zu fördern. Die letzte Hypothese dieses Erfolgsfaktors richtet sich abschließend nach einer allgemeinen und einer länderspezifischen Granularität speziell an einzelne Institutionen. Hierbei soll untersucht werden, ob diejenigen Institutionen, welchen ihrem Ruf nach eine höhere Leistungsfähigkeit in Bezug auf die Forschung nachgesagt wird, auch tatsächlich bessere Ergebnisse mit größerem Einfluss und Relevanz liefern, um durch

die daraus entstehenden Erkenntnisse dem Auftraggeber eine weitere Investitionsdimension zu eröffnen

Der Erfolgsfaktor Netzwerk, wie er durch das Lastenheft definiert wurde, wird im Rahmen unserer Analyse zwei getrennte Aspekte unterteilt. Zum einen sollen Netzwerkeffekte unter Forschungsinstitutionen untersucht werden. Des Weiteren wird analysiert, inwiefern sich Kooperationen zwischen Forschungseinrichtungen und Wirtschaftsunternehmen auf den Forschungserfolg auswirken. Durch diese differenzierte Analyse ist es möglich noch tiefere und präzisere Einblicke in die Materie zu erhalten und zukünftig gezielter die Forschung zu fördern. Im Rahmen der Analyse des Erfolgsfaktors Netzwerk unter Forschungseinrichtungen werden hierfür diese Hypothesen untersucht:

4. Der Forschungserfolg nimmt mit einer größeren Anzahl an beteiligten Institutionen zu

5. Länder mit einem größeren und aktiveren Kooperationsnetzwerk zwischen Forschungseinrichtungen haben einen größeren Forschungserfolg

Hypothese 4 untersucht, ähnlich wie Hypothese 1, den grundlegenden Einfluss des Erfolgsfaktors Kooperation auf den Forschungserfolg. Dabei wird untersucht, ob ein Zusammenhang zwischen der Anzahl an beteiligten Institutionen an Forschungsprojekten sich positiv auf die Ergebnisse auswirken. Damit wird auch hier die Basis für die Analyse dieses Erfolgsfaktors gelegt. Im Anschluss daran wird eine länderspezifische Analyse angesetzt. Hierbei soll untersucht werden, ob Länder, dessen Forschungseinrichtungen in einem größeren und frequenteren Ausmaß miteinander und auch länderübergreifend kooperieren, einen dementsprechenden Anstieg in ihrem Forschungserfolg verzeichnen können. Damit werden die Netzwerkeffekte in Bezug auf Forschungsinstitute voll abgedeckt.

Zusätzlich zur Analyse von Netzwerkeffekten unter Forschungseinrichtungen werden abschließend die Auswirkungen von Kooperationen der Forschung mit der Industrie untersucht. Hierbei werden folgende Hypothesen angebracht:

6. Länder mit häufigeren Kooperationen zwischen Forschungsinstituten und Wirtschaftsunternehmen verzeichnen einen größeren Forschungserfolg

7. Paper verzeichnen größere Forschungserfolge, wenn sie nicht nur von Forschungsinstitutionen, sondern in Kooperation mit Wirtschaftsunternehmen erarbeitet werden

Die allgemeinen Einblicke in die Netzwerkeffekte aus der Hypothese 4 werden nun um weitere Perspektiven speziell für die Kooperation mit Wirtschaftsunternehmen analysiert. Dabei wird an der länderspezifischen Sicht aus Hypothese 5 angeknüpft und auf die Zusammenarbeit von Forschungseinrichtungen und Unternehmen aus der Industrie erweitert beziehungsweise spezifiziert. Abschließend wird dann den bereits präsentierten Perspektiven erneut eine weitere Ebene mit einer präzisen Sicht auf einzelne Forschungsarbeiten hinzugefügt. Dabei soll untersucht werden, ob sich Forschungsarbeiten aus einer Kooperation zwischen Forschungseinrichtungen und Wirtschaftsunternehmen im Vergleich zu Forschungsarbeiten, welche lediglich aus Forschungseinrichtungen stammen, in ihrem Forschungserfolg abheben. Somit wird präzise untersucht, ob sich gezielte Förderungen in diesem Bereich als besonders effektiv herausstellen könnten.

Mit der Erweiterung und Spezifikation der gestellten Anforderungen im Bereich der zu untersuchenden Aspekte, decken wir hiermit eine möglichst diverse und tiefgründe Basis ab. In den nachfolgenden Kapiteln des Pflichtenhefts wird nun dargestellt, welche Methodiken und Datenquellen für die Analyse und Visualisierung der Hypothesen verwendet werden.

3. Datenquellen

Für die Durchführung der geplanten Analysen werden zwei zentrale und sich ergänzende Datenquellen herangezogen. OpenAlex wird hierbei als Metadatenbank für bibliometrische Informationen zu wissenschaftlichen Publikationen und Eurostat für finanzbezogene Forschungsdaten der EU-Mitgliedstaaten verwendet. Beide Quellen sind offen zugänglich, werden regelmäßig aktualisiert und ermöglichen eine flexible Nutzung im Rahmen des datengetriebenen Forschungsansatzes.

Die Datenbank OpenAlex dient als primäre Quelle für Informationen zu wissenschaftlichen Veröffentlichungen, Institutionen, Autoren und zugehörigen Metriken. Über die leistungsfähige OpenAlex-API wird ein direkter und dynamischer Zugriff auf aktuelle Daten ermöglicht. Die API erlaubt eine gezielte Filterung nach Themen, Publikationsjahr, Autorenzugehörigkeit, institutioneller Affiliation oder Ländern, was eine fein abgestimmte Auswahl der für die Hypothesen relevanten Paper erlaubt. Darüber hinaus liefert OpenAlex wichtige bibliometrische Kennzahlen wie Zitationsanzahl, Open-Access-Status und Field-Weighted Citation Impact (FWCI), welche als Indikatoren für den Erfolg von Forschung verwendet werden können. Durch die selektive API-Abfrage wird eine speicheroptimierte Datenverarbeitung ermöglicht, da nur relevante Datensätze abgerufen und verarbeitet werden müssen. Jedoch kann nicht davon ausgegangen werden, dass wirklich alle Paper auf dieser Plattform verfügbar und einsehbar sind. Nichtsdestotrotz handelt es sich hierbei gesichert um eine absolute Mehrheit dieser, weshalb hierbei dennoch nach den Anforderungen des Lastenhefts, eine repräsentative Stichprobe vorliegt. Diese wird zudem eingegrenzt, da nicht alle Parameter für jedes Paper zur Verfügung stehen. Diese werden demnach aus der Grundgesamt ausgeschlossen. Dennoch handelt es sich hierbei um eine ausreichende Datengrundlage für repräsentative Analysen.

Zur Ergänzung dieser publikationsbezogenen Datenbasis wird Eurostat als sekundäre Quelle genutzt. Eurostat stellt standardisierte und vergleichbare Daten zu den Forschungsausgaben der EU-Mitgliedstaaten bereit. Diese umfassen sowohl absolute Förderbeträge als auch normalisierte Werte, beispielsweise im Verhältnis zum Bruttoinlandsprodukt oder pro Kopf. Die Daten liegen auf unterschiedlichen Aggregationsebenen vor. Dabei werden diese national, sektoral oder nach Ausgabenträgern aufgeteilt und ermöglichen so eine differenzierte Analyse der finanziellen Rahmenbedingungen für Forschung innerhalb Europas. Es ist jedoch zu beachten, dass die Zahlen erst ab dem Jahr 2000 vollständig sind, weshalb dieses als zeitlicher Startpunkt der

Analyse definiert wird, welche bis in das Jahr 2024 reichen soll. Dieses Jahr wird hierbei als Endpunkt festgelegt, da zu dem Zeitpunkt des Projektes teilweise Daten für das aktuelle Jahr 2025 noch nicht vorliegen. Durch die Definition dieses Zeitraumes können wir die nicht-funktionalen Anforderungen der Aktualität und Repräsentativität der Daten sicherstellen. Zudem gilt Eurostat als eine der zentralen Datensammelpunkte Europas als äußerst zuverlässige Quelle, welche öffentlich zugänglich ist. Die Anforderungen in Bezug auf die Datenquelle des Lastenhefts werden hiermit also voll abgedeckt.

Die Kombination beider Quellen schafft eine solide empirische Basis für die geplanten Analysen. Während OpenAlex den wissenschaftlichen Output und dessen Wirkung abbildet, liefert Eurostat die finanziellen Inputgrößen. Dadurch wird es möglich, potenzielle Zusammenhänge zwischen Finanzierung, Kooperationen und Forschungserfolg in einem europäischen Vergleich datenbasiert zu untersuchen.

4. Analysekonzepte der Hypothesen

Für die Untersuchung der definierten Hypothesen werden jeweils unterschiedliche Methoden angewendet. Diese sind auf die jeweiligen Ziele zugeschnitten und sollen so optimale Ergebnisse fördern. Im Nachfolgenden wird Vorgehensweise für jede der sieben Hypothesen erläutert.

4.1. Hypothese 1

Bei der Analyse des Zusammenhangs der Menge an in die Forschung investiertem Geld und dem dadurch entstehenden Forschungserfolg stellt die investierte Summe des Geldes den Prädiktor dar, während der Forschungserfolg die Zielvariable verkörpert. Zum einen soll festgestellt werden, ob durch erhöhte Forschungsgelder auch die Anzahl an veröffentlichten wissenschaftlichen Papern zunimmt. Der quantitative Output der Forschung wird hierbei als ein positiver Faktor wahrgenommen, da mit einer höher frequenten Forschung und einer entsprechenden konsekutiven Verwendung und Weiterentwicklung der entstandenen Ergebnisse die Forschung im Allgemeinen schneller fortschreitet und hierdurch Innovationen fördert. Um aber auch die qualitativen Eigenschaften der Forschungsergebnisse nicht zu vernachlässigen, wird die Anzahl der Zitationen in absolutem sowie relativem Maß zu Papern verwandter Fachbereiche sowie der FWCI-Score untersucht, welcher die Relevanz des Papers quantifiziert. Diese Metriken treffen im Rahmen der Analyse eine Aussage darüber, wie wichtig und einflussreich die Ergebnisse der Forschung für diese waren und ob es sich hierbei um qualitativ hochwertige Arbeiten handelt. Als Datenquelle für die Finanzkennzahlen wird dabei Eurostat [1] herangezogen, während OpenAlex [2] die einzelnen Metriken und aggregierten Kennzahlen der wissenschaftlichen Arbeiten liefert.

Für die visuelle Darstellung der Ergebnisse dieser Untersuchung wird eine Zeitreihe sowie eine Regression verwendet. Hierbei soll zum einen der zeitliche Verlauf der verwendeten und verfügbaren Forschungsgelder im Zusammenhang mit dem Forschungserfolg dargestellt und zum anderen dieser Zusammenhang positive oder negative Verläufe der Regression verdeutlicht werden. Hierbei werden komplexe Diagramme angestrebt, welche zudem jeweils die diversen Metriken und geographischen Ursprünge der Forschungsarbeiten visuell vergleichen, um so tiefere Analysen zu ermöglichen und den Einfluss des Erfolgsfaktors Geld zu veranschaulichen.

4.2. Hypothese 2

Wie bereits bekannt wird im Rahmen der Hypothese 2 untersucht, wie effektiv die Investitionen in Forschung und Entwicklung in den einzelnen Mitgliedsstaaten der europäischen Union eingesetzt werden. Hierbei soll festgestellt werden, in welcher Relation der Grad an Forschungserfolg mit den investierten Geldern steht. Zum Beispiel wie viele Euro müssen in den einzelnen Staaten investiert werden, um ein Paper zu veröffentlichen oder 1000 Zitationen zu erwirken. Für einen Vergleich kann auf dieselben Metriken und damit Datenquellen zurückgegriffen werden, wie bei Hypothese 1, und werden dabei lediglich mit einer anderen Perspektive betrachtet. Hierfür werden die Metriken des Forschungserfolges jeweils in ein relatives Maß bezüglich der Mengen an Ressourcen gesetzt, um so eine aussagekräftige Verhältnisangabe zu erhalten

Für die Visualisierung dieser Zusammenhänge sollen für die einzelnen Metriken Heatmaps des europäischen Kontinents erzeugt werden, welche durch eine farbliche Codierung die räumliche Verteilung der jeweiligen Zustände in der Forschungseffizienz der Mitgliedsstaaten veranschaulichen. So werden geographische Unterschiede in der Effizienz der Nutzung von Forschungsressourcen in Relation zu den erarbeiteten Forschungsergebnissen oft leicht verständlich verarbeitet.

4.3. Hypothese 3

Die Untersuchung des tatsächlich messbaren Forschungserfolges von Elite-Institutionen wird ebenfalls anhand derselben Metriken und Datenquellen durchgeführt, welche auch schon für die beiden vorangegangenen Hypothesen für die Untersuchung des Forschungserfolges angebracht wurden. Die Datengrundlage wird dabei lediglich auf ausgewählte Einrichtungen beschränkt, deren Leistungsfähigkeit näher untersucht werden sollen. Die dafür angeführten Institutionen werden dabei aus Rankings wie dem Times Higher Education [3] oder Shanghai Ranking [4] entnommen, auf die EU-Mitgliedsstaaten eingegrenzt und entsprechend mit anderen Einrichtungen verglichen, welche nicht denselben Status besitzen.

Für die Visualisierung der Ergebnisse des Vergleichs dieser zwei Gruppen der Elite- und Nicht-Elite-Institutionen, wird ein Violin-Plot herangezogen. Dieser ermöglicht eine differenzierte Betrachtung der Verteilungen für die diversen Metriken pro Gruppe.

4.4. Hypothese 4

Durch die Analyse dieser These soll festgestellt werden, ob es sich positiv auf den Forschungserfolg auswirkt, wenn eine größere Anzahl an Institutionen an einem Projekt arbeiten, und somit beispielsweise ein größerer Pool an Expertise und Wissen auch zu besseren Ergebnissen führt. Die Metriken für das Messen des Forschungserfolges sind bereits aus den vorangegangenen Hypothesen bekannt. Die Anzahl an beteiligten Institutionen an einem Paper kann dabei ebenfalls aus der Metadatenbank OpenAlex als Parameter entnommen werden.

Die Ergebnisse können in einem zentralen Bubble-Chart effizient dargestellt. Durch die diversen Dimensionen, die durch Farbcodierung, Größe und Form der Bubbles dargestellt werden können, ermöglicht diese Visualisierungsform ein intuitives Verständnis des Zusammenhangs zwischen der Anzahl der beteiligten Institutionen und den verschiedenen Metriken des Forschungserfolges.

4.5. Hypothese 5

Für Hypothese 5 soll analysiert werden, ob Länder mit einem höheren Grad an institutioneller Kooperation zwischen Forschungseinrichtungen, einen höheren Forschungserfolg verzeichnen können. Die Zielvariable Forschungserfolg wird dabei durch die bereits etablierten Metriken dargestellt. Der Grad an Kooperation der einzelnen Länder wird durch diverse Metriken der Graphentheorie festgestellt. Darunter fallen die Degrees der einzelnen Knoten sowie deren Durchschnitte, die Betweenness und Closeness des Graphennetzwerkes. Hierbei wird gemessen, wie häufig Forschungseinrichtungen eines Landes mit anderen Forschungseinrichtungen desselben oder anderer Länder Forschungsprojekte durchführen und wie stark die dadurch entstehenden Netzwerke miteinander verknüpft sind. Dieser Grad an Kooperation wird dann in Relation zu dem zu verzeichnenden Forschungserfolg des Landes gesetzt, um so eine quantitative Aussage bezüglich der Hypothese treffen zu können.

Die Ergebnisse dieser Analysen sollen in entsprechenden Graphen dargestellt werden, welche durch die Abbildung dieser auf einer Karte von Europa veranschaulichen, welche Forschungseinrichtungen welcher Länder in einem größeren oder kleineren Maß miteinander kooperieren und gemeinsam Forschungsprojekte durchführen. Ein Knoten stellt dabei eine Institution dar, während eine Kante eine Kooperation zwischen zwei Institutionen darstellt. Durch die farbliche Unterscheidung der einzelnen Cluster,

sowie der Skalierung der jeweiligen Metriken durch unterschiedliche Ausprägungen der Kanten zwischen den einzelnen Knoten des abgebildeten Netzwerkes werden die Ergebnisse verständlich dargestellt. Die Koordinaten der Institutionen für die Darstellung auf der Europakarte werden Google Maps [5] entnommen.

4.6. Hypothese 6

Die Untersuchungsmethodik für die Netzwerkeffekte zwischen Forschungseinrichtungen und Wirtschaftsunternehmen ähnelt stark der von Hypothese 5. Der Unterschied liegt darin, dass nun zusätzlich die Unternehmen in das Netzwerk aufgenommen werden und lediglich die Verbindungen zwischen Forschungsinstituten und Wirtschaftsunternehmen visualisiert und gemessen werden. Demnach werden die Daten auf eine andere Art und Weise aggregiert, die Datenquellen sowie die Visualisierungsformen unterscheiden sich jedoch nicht.

4.7. Hypothese 7

Für Hypothese 7 soll untersucht werden, inwiefern Paper, an welchen sowohl Wirtschaftsunternehmen als auch Forschungseinrichtungen beteiligt waren, bessere Forschungserfolge verzeichnen als Paper, an welchen lediglich Forschungsinstitutionen gearbeitet haben. Die Metriken für den Forschungserfolg sowie die dafür notwendigen Datenquellen bleiben dabei erhalten.

Auch hier wird, wie bei der Untersuchung der Hypothese 3, welche ebenfalls eine binäre Gruppierung verfolgt, ein Violin-Plot verwendet. Dieser unterscheidet demnach den Forschungserfolg der Paper mit und ohne Beteiligung eines Wirtschaftsunternehmens.

5. Systemarchitektur und Projektstruktur

Die Umsetzung des Projekts erfolgt entlang einer strukturierten Datenpipeline, die alle Phasen der Analyse abbildet – von der Datenerhebung über die Auswertung bis hin zur finalen Ergebnispräsentation. Die Systemarchitektur ist dabei so gestaltet, dass ein klarer Arbeitsfluss sichergestellt wird und Verantwortlichkeiten im Team eindeutig zugeordnet sind.

Zu Beginn des Projekts steht die Erschließung und Überprüfung der relevanten Datenquellen. Hierbei werden vor allem die Eurostat-Datensätze sowie die API von OpenAlex hinsichtlich ihrer Verfügbarkeit, Struktur, Aktualität und Eignung für die jeweiligen

Analyseziele geprüft. Anschließend werden die Daten in einem mehrstufigen Prozess verarbeitet. Die zentrale Datenpipeline des Projekts gliedert sich in folgende Schritte:

1. Datenbeschaffung (via API-Zugriff oder Download)
2. Vorverarbeitung und Bereinigung
3. Analyse im Hinblick auf die formulierten Hypothesen
4. Visualisierung der Ergebnisse

Nach Abschluss der Analyse- und Visualisierungsphase folgt die Dokumentation der Ergebnisse im Projektbericht, in dem sowohl die Methodik als auch die gewonnenen Erkenntnisse systematisch dargestellt werden. Anschließend wird die multimediale Präsentation des Projekts erstellt, die als Videoformat die Data Story anschaulich und nachvollziehbar vermittelt. Zum Abschluss des Projekts erfolgt eine Retrospektive, in der das Vorgehen reflektiert und mögliche Verbesserungsansätze für zukünftige Vorhaben identifiziert werden.

Die Projektstruktur sieht eine klare Aufgabenverteilung innerhalb des Teams vor. Die Projektplanung und Koordination übernehmen Yanick und Carlo. Die Datenrecherche, einschließlich der technischen Überprüfung der Quellen, liegt bei David und Niklas. Die Formulierung der Hypothesen erfolgt gemeinschaftlich durch Niklas und Carlo, wobei die inhaltliche Bearbeitung nach Themenfeldern aufgeteilt ist. David und Niklas bearbeiten die Hypothesen im Bereich finanzielle Förderung, während Yanick und Carlo den Bereich Kooperation und Netzwerke übernehmen. Für die Erstellung des Projektberichts sowie der multimedialen Präsentation ist das gesamte Team gemeinsam verantwortlich, um eine konsistente und umfassende Ergebnisaufbereitung sicherzustellen.

Diese Aufteilung ermöglicht eine effektive Bearbeitung der Teilaufgaben, fördert die Verantwortung jedes Einzelnen und unterstützt eine stringente Umsetzung entlang der definierten Projektziele und Systemarchitektur.

6. Storytelling-Ansatz

Die Ergebnisse des Projekts werden in einer multimedialen Data Story aufbereitet, deren Ziel es ist, die gewonnenen Erkenntnisse nicht nur faktenbasiert, sondern auch verständlich, nachvollziehbar und in einem größeren gesellschaftlichen Kontext zu vermitteln. Der Storytelling-Ansatz orientiert sich dabei an einem klar strukturierten dramaturgischen Aufbau, der sich in drei Hauptphasen gliedert: Einleitung, Analyse und Fazit.

Zu Beginn der Data Story wird eine motivierende Einleitung gegeben, in der die Relevanz des Themas im europäischen Kontext verdeutlicht wird. Der Fokus liegt hierbei auf der zentralen Frage, warum Forschung für die langfristige Zukunftsfähigkeit der Europäischen Union von entscheidender Bedeutung ist. Aspekte wie Innovation, Nachhaltigkeit, wirtschaftliche Wettbewerbsfähigkeit und gesellschaftlicher Fortschritt werden kurz umrissen, um die Notwendigkeit einer fundierten Analyse von Erfolgsfaktoren in der europäischen Forschungslandschaft zu begründen.

Im Anschluss folgt der Hauptteil der Präsentation, in dem die drei im Projekt untersuchten Hypothesen nacheinander vorgestellt werden. Für jede Hypothese wird zunächst in kurzer Form erläutert, welcher Aspekt des Forschungssystems analysiert wird, warum diese Fragestellung relevant ist und welche Bedeutung sie für zukünftige strategische Entscheidungen der EU haben kann. Daraufhin werden die zentralen Ergebnisse der jeweiligen Analyse präsentiert und visuell veranschaulicht. Dieser Teil bildet den analytischen Kern der Data Story und soll dem Publikum ermöglichen, die Zusammenhänge zwischen Förderstruktur, Kooperationen und Forschungserfolg nachvollziehbar zu erkennen.

Den Abschluss der Präsentation bildet ein kompaktes Fazit, in dem die übergeordneten Erkenntnisse aus den Einzelanalysen zusammengeführt werden. Dabei wird aufgezeigt, welche strukturellen Merkmale Forschung in Europa begünstigen, und wie diese für die Gestaltung zukünftiger Forschungsförderung auf europäischer Ebene genutzt werden könnten – ohne jedoch normative Handlungsempfehlungen auszusprechen.

Die Data Story verfolgt damit das Ziel, komplexe datenbasierte Inhalte in eine ansprechende und logisch aufgebaute Erzählform zu überführen, die sowohl informativ als auch zugänglich für ein breites Publikum ist. Durch die Verbindung von analytischer

Tiefe und narrativer Klarheit soll ein Beitrag zur evidenzbasierten Auseinandersetzung mit der Entwicklung europäischer Forschung geleistet werden.

7. Technologien und Tools

Für die Umsetzung des Projekts kommen verschiedene spezialisierte Technologien und Werkzeuge zum Einsatz, die sowohl die Datenbeschaffung und Analyse als auch die Visualisierung und finale Aufbereitung unterstützen. Dabei wird insbesondere auf bewährte, frei verfügbare Open-Source-Lösungen gesetzt, die eine hohe Flexibilität, Reproduzierbarkeit und Erweiterbarkeit gewährleisten.

Die Daten werden über die offizielle Python API von OpenAlex bezogen. Diese ermöglicht einen direkten, flexiblen und tagesaktuellen Zugriff auf umfassende bibliometrische Informationen zu wissenschaftlichen Publikationen, Institutionen und Autoren. Durch gezielte API-Abfragen können genau jene Daten extrahiert werden, die für die jeweiligen Thesen und Analysen relevant sind.

Zur Datenverarbeitung und -analyse wird primär die Programmiersprache Python eingesetzt, insbesondere unter Verwendung der Bibliothek Pandas, die eine leistungsfähige Bearbeitung, Aggregation und Filterung strukturierter Datensätze erlaubt. Für weiterführende statistische Auswertungen und Metrikberechnungen bildet Pandas die zentrale Grundlage der Analyseumgebung.

Im Rahmen der Untersuchung von Kooperationsstrukturen kommt zusätzlich die Bibliothek NetworkX zum Einsatz. Mit ihr lassen sich institutionelle und sektorübergreifende Netzwerke modellieren, analysieren und visualisieren. Dadurch können Verbindungen und Muster in den Kooperationsbeziehungen sichtbar gemacht und hinsichtlich ihrer potenziellen Wirkung auf den Forschungserfolg interpretiert werden.

Zur räumlichen Verortung und Darstellung regionaler Muster wird GeoPandas in Kombination mit Matplotlib verwendet. Diese Bibliotheken ermöglichen die Erstellung von Kartendarstellungen, etwa zur Visualisierung von Publikationsverteilungen oder grenzüberschreitenden Forschungskooperationen auf Landes- oder Institutionsebene.

Für die Organisation des Quellcodes und zur Sicherstellung der Nachvollziehbarkeit im Team wird GitHub als Plattform für Versionierung, Code-Review und kollaboratives Arbeiten genutzt. Änderungen am Code werden dokumentiert, getestet und strukturiert verwaltet, sodass jederzeit der aktuelle Entwicklungsstand reproduzierbar ist.

Zur Erstellung der multimedialen Abschlusspräsentation wird PowerPoint als Basis verwendet. Die Inhalte der Data Story werden dabei zunächst in visuell unterstützter Form strukturiert, anschließend mit ergänzendem Sprechertext vertont und abschließend als Videoexport für die Einreichung und Präsentation aufbereitet.

Die Kombination dieser Technologien stellt sicher, dass das Projekt sowohl technisch fundiert, methodisch nachvollziehbar als auch visuell und kommunikativ überzeugend umgesetzt werden kann.

8. Qualitätssicherung

Um eine hohe inhaltliche, technische und visuelle Qualität im Projekt sicherzustellen, werden im Verlauf der Umsetzung gezielte Maßnahmen zur Qualitätssicherung ergriffen. Diese betreffen sowohl den Umgang mit Datenquellen, die technische Umsetzung der Analysen als auch die Präsentation der Ergebnisse.

Ein zentraler Aspekt ist die differenzierte Auswahl und Prüfung der Datenquellen. Es wird darauf geachtet, dass ausschließlich öffentlich zugängliche, dokumentierte und als zuverlässig geltende Datenquellen (z. B. OpenAlex, CORDIS, Eurostat) verwendet werden. Die Relevanz und Vollständigkeit der jeweiligen Datensätze wird im Vorfeld jeder Analyse geprüft, und bei identifizierten Schwächen werden Alternativquellen herangezogen oder Einschränkungen dokumentiert.

Zur Sicherstellung der technischen Nachvollziehbarkeit und Reproduzierbarkeit wird die komplette Projektumsetzung in einer strukturierten und versionierten Codebasis organisiert. Hierfür wird GitHub als zentrales Werkzeug für Versionskontrolle, Kollaboration und Änderungsverfolgung eingesetzt. Alle Codebestandteile – von der Datenbeschaffung über die Vorverarbeitung bis zur Visualisierung – werden nachvollziehbar dokumentiert. Zur Qualitätssicherung des Codes werden regelmäßige Code-Reviews im Team durchgeführt. Dabei wird insbesondere auf funktionale Korrektheit, Lesbarkeit, Modularität sowie Wiederverwendbarkeit geachtet.

Auch die Gestaltung und Darstellung der Analyseergebnisse unterliegt definierten Qualitätskriterien. Sämtliche Grafiken und Diagramme werden als Vektorgrafiken exportiert, um auch bei Skalierung eine verlustfreie Darstellung zu ermöglichen. Farbwahl, Achsenskalierung, Beschriftungen und Legenden werden einheitlich und barrierearm gestaltet, um eine hohe Verständlichkeit sicherzustellen. Für besonders

komplexe Darstellungen (z. B. Netzwerke) werden Prototypen intern getestet und ggf. angepasst, um Überfrachtung oder Fehlinterpretationen zu vermeiden.

Zur Wahrung der inhaltlichen Konsistenz wird auf eine durchgängige Dokumentation aller Analyseentscheidungen, Filterkriterien und genutzten Metriken geachtet. Die verwendeten Definitionen, Variablen und Berechnungen werden in der Methodendokumentation festgehalten, sodass eine Reproduktion und Weiterverwendung der Ergebnisse durch Dritte möglich ist.

Die Einhaltung der genannten Maßnahmen soll gewährleisten, dass die erarbeiteten Ergebnisse sowohl technisch als auch inhaltlich fundiert, korrekt und überzeugend kommunizierbar sind. Die Qualitätssicherung bildet damit eine tragende Säule für die wissenschaftliche Seriosität und langfristige Nutzbarkeit des Projekts.

9. Risiken und Gegenmaßnahmen

Im Rahmen der Projektumsetzung können verschiedene Risiken auftreten, die den zeitlichen Ablauf, die Datenqualität oder die Aussagekraft der Ergebnisse beeinträchtigen könnten. Ziel dieses Kapitels ist es, potenzielle Risiken frühzeitig zu identifizieren und geeignete Gegenmaßnahmen zu definieren, um die Projektziele dennoch zuverlässig erreichen zu können.

Ein zentrales Risiko liegt in der Verfügbarkeit und Qualität der verwendeten Datenquellen. Insbesondere bei der Nutzung von offenen Schnittstellen wie der OpenAlex-API kann es zu temporären Ausfällen, Rate Limits oder strukturellen Änderungen kommen. Um diesem Risiko zu begegnen, wird frühzeitig ein vollständiger Datenabzug als Backup eingeplant. Zudem werden alternative Quellen (z. B. CORDIS oder Dimensions) evaluiert, um im Bedarfsfall auf andere Datenbasis ausweichen zu können.

Ein weiteres Risiko betrifft die Komplexität der Netzwerkanalyse. Bei sehr großen oder unvollständigen Netzwerkstrukturen kann es zu Schwierigkeiten bei der Interpretation, Performance oder Visualisierung kommen. Um dies zu vermeiden, werden geeignete Filter- und Clustering-Verfahren eingesetzt, um die Visualisierungen auf eine interpretierbare Größe zu begrenzen. Zusätzlich erfolgt eine iterative Validierung der Netzwerkanalyseergebnisse im Team.

Auch im Bereich der Visualisierung besteht das Risiko, dass die erzeugten Diagramme für Außenstehende schwer verständlich oder überladen sind. Zur Vermeidung dessen werden Visualisierungen gezielt nach narrativen Kriterien gestaltet, auf Kernaussagen

reduziert und vorab intern getestet. Technisch wird zudem auf die Verwendung skalierbarer Vektorgrafiken geachtet, um die Qualität auch in multimedialen Formaten zu sichern.

Organisatorisch besteht das Risiko von zeitlichen Engpässen im Projektverlauf, insbesondere bei der Finalisierung der multimedialen Präsentation. Um dieses Risiko zu minimieren, wird ein klarer Zeitplan mit Meilensteinen und internen Zwischenterminen verfolgt. Aufgaben werden frühzeitig aufgeteilt und die finale Präsentation parallel zur Ergebnisdokumentation vorbereitet, um Pufferzeiten sicherzustellen.

Schließlich wird auch der Zusammenarbeit im Team besondere Beachtung geschenkt. Missverständnisse in der Aufgabenverteilung oder der technischen Umsetzung können den Fortschritt behindern. Durch regelmäßige Absprachen, ein zentrales Repository und eine transparente Dokumentation der Analyseentscheidungen wird die Zusammenarbeit aktiv abgesichert.

Insgesamt sind die identifizierten Risiken beherrschbar, sofern die genannten Gegenmaßnahmen konsequent umgesetzt werden. Die Qualitätssicherung, flexible Datenstrategien und ein strukturiertes Projektmanagement bilden dabei die zentralen Instrumente zur Risikominimierung im Projekt.

10. Bemerkungen

Derzeit liegen keine ergänzenden Bemerkungen vor. Sollte es im weiteren Projektverlauf zu Anmerkungen kommen, die nicht den bestehenden Kapiteln zugeordnet werden können, werden diese an dieser Stelle dokumentiert.

11. Finale Freigabe

Dieses Dokument beinhaltet in der Version 1.0 die genauen Hypothesen, Datenquellen, Untersuchungs- und Visualisierungsmethoden sowie die zu beachtenden Aspekte bezüglich der Qualitätssicherung und der Projektrisiken. Das Projekt soll planmäßig wie hier beschrieben durchgeführt werden.

Mannheim, 23.05.2025

Ort, Datum

C. Rinderer

Unterschrift Projektleiter
(C. Rinderer)

12. Quellenverzeichnis

[1] Eurostat, URL: <https://ec.europa.eu/eurostat/de/data/database>, zuletzt abgerufen: 23.05.2025

[2] OpenAlex, URL: <https://openalex.org/>, zuletzt abgerufen: 23.05.2025

[3] Times Higher Education, URL: <https://www.timeshighereducation.com/>, zuletzt abgerufen: 23.05.2025

[4] Shanghai Ranking, URL: <https://www.shanghairanking.com/>, zuletzt abgerufen: 23.05.2025

[5] Google Maps, URL: <https://www.google.com/maps/>, zuletzt abgerufen: 23.05.2025