

# Identifikation von Kosten- faktoren und Erarbeitung von Kostenmodellen (R 1.1.1)

**Version** 28.03.2014

**Arbeitspaket** 1.1

**verantwortlicher Partner** IDS

## TextGrid

Virtuelle Forschungsumgebung für die Geisteswissenschaften



GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

**Projekt:** TextGrid – Institutionalisierung einer Virtuellen Forschungsumgebung in den Geisteswissenschaften

**BMBF Förderkennzeichen:** 01UG1203A

**Laufzeit:** Juni 2012 bis Mai 2015

**Dokumentstatus:** Final

**Verfügbarkeit:** Öffentlich

**Autoren:**

Fiedler, Norman; IDS

Funk, Stefan; DAASI

Gietz, Peter; DAASI

Küster, Marc W.; FH Worms

Rapp, Andrea, TU Darmstadt

Söring, Sibylle; SUB Göttingen

Vitt, Thorsten; Univ. Würzburg

Wieder, Philipp; GWDG

Witt, Andreas; IDS

**Revisionsverlauf:**

Datum	Autor	Kommentare
25.11.2013	N. Fiedler	Entwurf
22.01.2014	A. Werthmann	Korrekturen
13.03.2014	Alle	Redaktion
28.03.2014	N. Fiedler	Finalisierung

# Inhaltsverzeichnis:

<b>1. Einleitung .....</b>	<b>4</b>
<b>2. Stand der Kostenermittlung .....</b>	<b>5</b>
2.1. Methodik.....	5
2.2. Kostenarten .....	7
<b>3. Kostenkalkulation für TextGrid .....</b>	<b>12</b>
3.1. Kostenkategorie Infrastruktur .....	12
3.2. Kostenkategorie Betrieb/Softwarepflege Lab .....	14
3.3. Betrieb Softwarepflege Rep.....	16
3.4. Nutzerbetreuung.....	17
3.5. Organisatorischer Betrieb .....	19
<b>4. Fazit .....</b>	<b>20</b>
<b>5. Literatur.....</b>	<b>21</b>

# 1. Einleitung

Die zunehmende Tendenz auch geisteswissenschaftlicher Forschung hin zu einer empirisch gestützten Wissenschaft und mithin das Anwachsen von digitalen Forschungsdaten optimiert nicht nur die Verifizierbarkeit von Forschungsergebnissen, es werfen sich vielmehr auch Herausforderungen jenseits ihrer technischen und organisatorischen Handhabung auf. Angesichts knapp bemessener Haushalte sollte jede Konzeption nachhaltigen Forschungsdatenmanagements die Fragen nach der Finanzierbarkeit und der Kostenverteilung des stetig zunehmenden Datenstroms mithdenken. Hierbei reicht jedoch eine kurzfristige, initial ausgelegte Finanzierung nicht mehr aus, begreift man die Bestandserhaltung des Datenpools als langfristige Aufgabe der Wahrung von einmalig gewonnenen Erkenntnissen sowie der Vermeidung von redundanten Neuerhebungen von Forschungsdaten im Verlustfalle [Rad2, 18ff.].

Virtuelle Forschungsumgebungen (VFU) sollen Wissenschaftlern als integrierte Werkzeugplattform zur Erzeugung, Bearbeitung, Ablage und Nachnutzung digitaler Forschungsdaten dienen. Insbesondere der Aspekt ihrer Nachnutzung zum Zwecke der Replizierbarkeit und Falsifizierbarkeit von Forschungsdaten und den mit ihrer Hilfe generierten Forschungsergebnissen erfordert nicht nur technische Flexibilität und Entwicklungspotenzial mit Blick auf sich ändernde Verfahren und Formate, sondern auch einen nachhaltigen Betrieb von VFUen, so dass die empirische Basis der Forschung mittelfristig nicht wegbreicht. Die genannten Anforderungen von Nutzern an VFUen lassen sich auf die Komponenten einer umfassenden Nachhaltigkeitsstrategie abbilden. Seit Längerem nehmen Förderer das Kriterium der Verstetigung der beantragten Vorhaben in die Formalitäten ihre Ausschreibungen auf - eine Mindestvorhaltezeit von Daten und Werkzeugen von 10 Jahren gilt bei der Deutschen Forschungsgemeinschaft (DFG). Solche Vorgaben sind jedoch gerade für Hochschulen und hier insbesondere für die kleinen geistes-, sozial- und kulturwissenschaftlichen Disziplinen kaum zu bewerkstelligen bzw. werden kaum berücksichtigt, obgleich sie bisweilen auch von der Möglichkeit der zusätzlichen Beantragung von Geldern zum Zwecke der Nachnutzbarkeit genehmigter Vorhaben flankiert werden [Beucke, 26]. Hieraus ergibt sich die Notwendigkeit eines Wissens um die Befähigung der Wissenschaft zur Kalkulation der Kosten des Forschungsdatenmanagements bereits vor der Erzeugung der Daten und mit der Antragstellung. Perspektivisch kommen Antragsteller also nicht um die genaue Einschätzung der Kosten der Nachhaltigkeit herum, wollen sie Fördergelder in seriös beziffertem Umfang einwerben [DFG, 21f.; Rad1, 6; GESIS, 18].

So hat auch das Projekt TextGrid die Erarbeitung eines solchen systematisierten Vorgehens in seinen Arbeitsplan aufgenommen. Neben Maßnahmen der technischen und fachwissenschaftlichen Nachhaltigkeit, welche die technische Les- und Interpretierbarkeit der Forschungsdaten sowie die Verankerung des Wissens um den Umgang mit VFUen in der wissenschaftlichen Nutzergemeinde absichern, obliegt es der organisatorischen Nachhaltigkeit, ein tragfähiges organisatorisches Gerüst und vor allem eine stabile Finanzierung zu etablieren.

Derzeit und in der Vergangenheit erfolgte die Finanzierung des Aufbaus und – inzwischen in der dritten Förderphase – des testweisen Echtbetriebs von TextGrid über eine zeitlich befristete projektgestützte Förderung. Um eine mittel- bis langfristig stabile (als Richtwert werden zunächst 10 Jahre angenommen) Finanzierung zu realisieren, sollte eine jede Einzelförderung in eine institutionelle Förderung übergehen [WissGrid, 23]. Die wesentlichen Akteure, denen

eine Finanzierung der VFU perspektivisch zukommen wird, sind hierbei natürlich an der Aufschlüsselung der beim Betrieb entstehenden Kosten interessiert, sodass für die weitere Entwicklung eines Finanzierungsmodells die Kostenermittlung im Rahmen eines Kostenmodells erfolgen muss.

Der Versuch, dies für eine VFU in den genannten Fachdisziplinen zu unternehmen, stellt eine erste Annäherung dar, zumal es einige Unwägbarkeiten zu berücksichtigen gilt. So bemerkte die Kommission zur Zukunft der Informationsinfrastrukturen (KII) der Leibniz-Gemeinschaft speziell zu den Kosten einer VFU:

*„Kostenermittlungen zur Absicherung der mittel und langfristigen Verfügbarkeit leiden an der noch geringen Etabliertheit solcher Dienste und in Folgedessen an mangelnden internationalen Erfahrungswerten gerade auch angesichts der großen Heterogenität des Materials, des hohen Volumens und der Anforderungen“* [KII; 34].

Nimmt man vor diesem Hintergrund den Grundwert als gegeben an, dass insbesondere die Resultate (öffentlicht geförderter) wissenschaftlicher Forschung zumindest für wissenschaftlicher Nutzer unter Wahrung rechtlicher Normen uneingeschränkt, langfristig und frei zur Verfügung stehen sollten, verbieten sich für eine Finanzierungslösung fast kostenpflichtige Umlagemodele und die Frage nach der Finanzierbarkeit stellt sich umso dringender [WiSta, 1239]. Ist weiterhin beabsichtigt, die Finanzierung des Forschungsdatenmanagements auf eine stabile Grundlage zu stellen, indem verlässliche Partner für eine dauerhafte Kostenübernahme gewonnen werden sollen, gilt es als ersten Schritt, Arten und Höhe der Kosten vorab zu kalkulieren [Beucke, 26], „*denn niemand wird bereit sein, für Kosten geradezustehen, deren Höhe völlig offen ist*“ [Rad1, 3].

## 2. Stand der Kostenermittlung

### 2.1. Methodik

Zwar existieren zahlreiche Untersuchungen zu Kostenermittlungen rund um den Umgang mit Forschungsdaten, doch konzentrieren sich diese fast ausschließlich auf den Themenkomplex des Lebenszyklus von Forschungsdaten. Im Fokus stehen die Erhebung (Acquisition), die Überführung in ein Archiv oder Repozitorium (Ingest), die Langzeitarchivierung (LZA) sowie die Verfügbarmachung ([Open] Access) und Verfügbarhaltung bzw. Nachnutzung von Forschungsdaten [Beargie, 24ff.]. Adressat und Interessent solcher Untersuchungen sind meist standortgebundene wissenschaftliche Organisationen, Bibliotheken und Archive, welche die mit zunehmender Datenflut und Datenverlust verbundenen Probleme zu adressieren beabsichtigen. Eine dedizierte Betrachtung von VFUen fand bisher vielleicht aufgrund der von der KII konstatierten mangelnden Erfahrungswerte noch nicht statt.

Lediglich das Projekt Radieschen<sup>1</sup> nimmt die VFU TextGrid kurz auf und betrachtet den finanziellen Aufwand für das Vorhaben auf Basis der bisherigen für Förderphase III, welche eine Test- und Übergangsphase für den Realbetrieb darstellt, bereitgestellten Projektmittel.

---

<sup>1</sup> <http://www.forschungsdaten.org/index.php/Radieschen>

Aufgestellt werden Aufbau-, Entwicklung sowie Produktionskosten, die den jährlichen Bedarf auf € 926.062 beziffern [Rad1, 15]. Nichtsdestotrotz soll ein rascher Überblick über bisherige Ansätze zum Forschungsdatenmanagement nicht nur deswegen gegeben werden, weil die Speicherung und Zugänglichmachung von Daten Kernbestandteil jeder VFU ist, sondern um methodische Entlehnungen für die weitere Analyse gewinnen zu können.

Grundsätzlich gilt es vorauszuschicken, dass sich eine ausschließliche Archivierung von Forschungsdaten hinsichtlich ihres finanziellen Aufwandes gegenüber den Investitionen in die Forschungsarbeit, die zur Erzeugung dieser Daten führt, vergleichsweise gering ausnehmen. Britische Untersuchungen haben gezeigt, dass die Kosten des Forschungsdatenmanagements in den Gesamtaufwendungen für die Forschung lediglich bis zu 1,5% ausmachen, aber die Kosten für reine Publikationsarchive (auch elektronische Dokumente) deutlich übertreffen [Rad2, 3]. Wie ist es zu solchen Erkenntnissen gekommen?

Das Projekt *Radieschen* hat in einer Interviewreihe wissenschaftliche Institutionen nach der Kostenverteilung in ihrem Forschungsdatenmanagement befragt, ohne jedoch einzelne Posten mit zugehörigen Geldbeträgen gesondert zu eruieren. Infolgedessen wurde auch kein Kostenmodell erarbeitet wie es Ziel dieses Berichts sein soll. Vielmehr konzentrierten sich die Betrachtungen in *Radieschen* auf den, wie schon durch andere Studien postuliert und hier alsbald bestätigt, größten Block der Aufwendungen: die Personalkosten. Die Bearbeiter weisen nicht zu Unrecht darauf hin, dass mit diesem Ansatz kaum Geldbeträge in Erfahrung zu bringen, jedoch Informationen zu in Personenmonaten gerechnete Personalstellen vergleichsweise leicht einzuholen sind. Der Vorteil einer auf Personalstellen ausgerichteten Befragung besteht hingegen in der Möglichkeit der Aufschlüsselung von finanziellen Aufwänden und ihrer Zuordnung zu einzelnen Zuständigkeiten und mithin von Kostenfaktoren innerhalb des Forschungsdatenmanagements [Rad2, 4]. Um zu den für einzelne Kostenfaktoren anfallenden Aufwänden zu gelangen, wurden schließlich sämtliche Stellenzahlen zu einem Arbeitsschritt des Archivierungs- und Bereitstellungsprozesses kumuliert, um anschließend einen an der Zahl der Befragten orientierten Durchschnittswert für jeden Posten zu gewinnen [Rad2, 7]

Methodisch zeigt *Radieschen* eine Annäherung an eine Kostenermittlung mittels zweier Ansätze auf.

*„Die Kostenrechnung kann deduktiv oder induktiv erfolgen. Bei der deduktiven Methode wird ausgehend von den Gesamtkosten versucht, durch Abgrenzung zu Einzelkosten zu kommen. Der induktive Ansatz geht von den einzelnen Arbeitsschritten aus. Die Einzelschritte werden zu Teilprozessen und schließlich zu Hauptprozessen verdichtet. Die Kosten für die Einzelschritte müssen bekannt sein oder geschätzt werden, damit die Kosten für die Teil- und Hauptprozesse berechnet werden können. Umgekehrt müssen beim deduktiven Ansatz die prozentualen Anteile bekannt sein, um die Abgrenzung durchführen zu können.“* [Rad1, 9]

Grundsätzlich ist hierbei davon auszugehen, dass sich die Kosten des Forschungsdatenmanagements – insbesondere die wichtigen Investitionen in Personal und Hardware – langfristig betrachtet nicht signifikant anders entwickeln als andere Investitionen in Forschungsinfrastrukturen. Demnach sinken die Kosten für Hardware im Laufe der Betriebszeit von sehr hohen Investitionsausgaben bis hin zu Kosten für Wartung und Pflege spürbar, während Gehälter der betreuenden Wissenschaftler kontinuierlich aber langsam anwachsen. Für ein nachhaltiges Finanzierungskonzept ist daher zu beachten, dass ein einmalig erstelltes Kostenmodell

nicht über die gesamte Betriebszeit des Dienstes unverändert beibehalten werden kann. Insbesondere Ausschläge in der Nutzungsintensivität, Änderungen von Anforderungen an das Leistungsprofil sowie Art und Umfang der zugrundegelegten Technologien und Daten führen zu Schwankungen in den Kosten des Forschungsdatenmanagements [Rad1, 9f.].

Das von der DFG geförderte Göttinger Vorhaben „Kooperative Langzeitarchivierung für Wissenschaftsstandorte“ (*KoLaWiss*)<sup>2</sup> geht für die Entwicklung seines Kostenmodells von der Grundannahme aus, dass die LZA von einem Referenzstandort aus erbracht und weitgehend dort genutzt wird. Eine auf mehrere Standorte redundant verteilte Archivierung erfordere hingegen ein Kostenmodell, das auf dem Dienstleistungsprinzip beruhe und erbrachte Leistungen ausgleiche, um die finanzielle Stabilität des Konstrukts nachhaltig zu gewährleisten [KLW, 4]. Im weiteren Vorgehen müssen nach diesem Ansatz zunächst Kostenarten identifiziert und diese in einem zweiten Schritt bei Dienstleistern erhobenen Tarifen zugeordnet werden. Personalkosten ergeben sich hingegen über eine eingehende Analyse der dem Langzeitarchiv zugrundeliegenden Organisationsstruktur [KLW, 5f.].

## 2.2. Kostenarten

Zur Kalkulation der Aufwände, die beim Umgang mit Forschungsdaten entstehen, ist der entscheidende Schritt die Faktoren zu identifizieren, die Kosten verursachen. Verschiedene Studien haben solche Faktoren – in den meisten Fällen Kostenarten genannt – ausgemacht.

*Radieschen* bestimmt zunächst grobe Tätigkeitsbereiche – als Domänen bezeichnet – die sich querschnittlich über zu den Kostenarten verhalten und deren Funktionsweise und Kostenhöhe bestimmen. Es wird hierbei zwischen der privaten Domäne des Wissenschaftlers (Privatarchive), der Gruppendomäne (Projekt- oder Arbeitsgruppenbezogene Archive), der Öffentlichen Domäne (institutionelle Archive) und der Zugangsdomäne (Datenzugriff; in privater bis öffentlicher Domäne für jeweilige Nutzergruppen inhärent, für externe Zugriffe zusätzlich anzusetzen) unterschieden. Basierend auf den Ergebnissen des Projekts *nestor*<sup>3</sup> wurden diesen Domänen Kostenarten zugeordnet. Identifiziert wurden zu diesem Zweck

- Initiale Kosten (Vorbereitung und Planung der Archivierung, Sichtung der Ressourcen),
- Beschaffungskosten (Hardware, Lizizenzen, Schaffung von Personalstellen),
- Entwicklungskosten,
- und Betriebskosten (Personal, Weiterbildung, Storage, Systembetrieb und Backup, infrastrukturelle und leistungsbezogene Gebühren und Mieten, Expansion und Ersatz des Gerätbestandes) [Rad1, 5, Tabelle 1].

Zu den wesentlichen Erkenntnissen dieser Zuordnung gehört, dass mitarbeiterbezogene Kosten (Gehälter, Reisemittel, Weiterbildung) ebenso wie Hardwarekosten zumeist in der Gruppen- als auch in der öffentlichen Domäne anfallen. Andererseits werden durch infrastrukturelle Anschaffung und Betrieb entstehende Kosten (Gebäudekosten, Strom, Kühlung, Netzwerke) kaum von den Archivorganisationen selbst, sondern zumeist von den hinter ihnen stehenden

<sup>2</sup> <http://www.gwdg.de/index.php?id=213&L=aknhodasfdcjmjy>

<sup>3</sup> <http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch.pdf>

den Trägereinrichtungen getragen. Entwicklungskosten fallen indes gleichermaßen bei Gruppen bis Zugangsdomänen an. [Rad1, 4]

Die von der Leibniz-Gemeinschaft eingesetzte *KII*<sup>4</sup> hat ein nationales Gesamtkonzept für Konzeption und Betrieb solcher Einrichtungen erarbeitet, im Rahmen dessen Kosten zu einzelnen Handlungsfeldern beleuchtet werden. An dieser Stelle sei die Wiedergabe der Ergebnisse auf die Themenkomplexe LZA, VFU, Open Access und Forschungsdaten beschränkt. Eine Ressourcenabschätzung der LZA wird dahingehend problematisiert, als komplexe Kostenarten zu berücksichtigen sind. Hierunter fallen zunächst die an den Nutzeranforderungen ausgerichteten Dienste, die sich, nachhaltig aufgestellt, synergetisch an bestehende und im Aufbau befindliche Vorhaben ähnlicher Ausrichtung zu koppeln haben. Des Weiteren sind Anschaffungs- und Betriebskosten von Hard- und Software ebenso abzudecken wie der Unterhalt abstützender organisatorischer Strukturen samt des dort versammelten Personals. Zu guter Letzt ist im Sinne der fachwissenschaftlichen Nachhaltigkeit für die Weiterbildung, Qualifizierung und Vernetzung der Mitarbeiter Sorge zu tragen. Zur nachhaltigen Organisation von LZA-Diensten liegen der Studie keine ausreichenden Erfahrungswerte vor, sodass kaum Aussagen über die langfristig anfallenden Kosten getätigt werden können. In einer groben Schätzung aus dem Bereich der Nationallizenzen mit der Archivierung vergleichsweise kleiner Objekttypen über etwa 50 TeraByte (50 Mio. Artikel zu je 1 MB) wird mit € 400.000 jährlich gerechnet, legt man eine einfach redundante Datenhaltung samt Backup zugrunde. Ein ebenso grober, auf Erfahrung basierter Richtwert stellt die Kalkulation der Kosten für Service und Bereitstellung der Daten vom 1,7fachen der Speicherkosten dar. Hostingkosten summieren sich demnach auf etwa € 1,1 Mio. jährlich. Einschränkend gilt jedoch:

*„Für die Langzeitarchivierung gelten andere Bedingungen (andere Speichercharakteristik, Mehrfachkopien, verteilte Datenhaltung, Überwachung, aufwändiger Service- und Betriebsprozesse, dedizierte Normalisierung- und Ingest-Abläufe, laufende intellektuelle Überwachungsprozesse (Risiko-Management), gezielte Migrationsprozesse usw.) Hier ist über die eigentlichen Speicherkosten hinaus (diese sind hier gleich angesetzt, allerdings mit anderen Kostenanteilen) ein Faktor von 2,2 zu den Speicherkosten anzusetzen, in der Summe ergeben sich 1,3 Mio. EUR pro Jahr. Insgesamt würde nach dieser groben Schätzung die mittel- bis langfristige Sicherung des Zugriffs und der Verfügbarkeit von „Nationallizenzen“ Kosten in Höhe von 2,4 Mio. EUR pro Jahr verursachen.“* [KII, 35]

Die Kosten für VFUen sind, wie bereits angeführt, aufgrund mangelnder Erfahrungswerte jenseits kurzfristiger Projektförderung kaum abzuschätzen. Letztere entfallen zumeist auf die Aufbauphase der VFU, während die Betriebsphase eine langfristige Investition in die Pflege der bestehenden und die Addition neuer Funktionalitäten und Komponenten mit sich bringt. Für die Aufbauphase wird auf Durchschnittswerte der Projektförderer zurückgegriffen – insbesondere wird eine Querschnittziehung der BMBF-geförderten D-Grid-Initiative versucht. Bei der DFG liegen die Größenwerte zwischen € 300.000 bis € 500.000 pro Fördermaßnahme. Besonders nutzer- und datenintensive Forschungsumgebungen der Naturwissenschaften kommen demnach auf einen Mittelwert von € 5 Mio. für die initiale Etablierung der Dienste. Für die hier eigentlich relevante Betriebsphase liegen hingegen keine ausreichenden Daten vor [KII, 40].

---

<sup>4</sup> <http://www.leibniz-gemeinschaft.de/infrastrukturen/kii/>

Auch im Sektor des Open Access bewegt sich die *KII* recht vorsichtig. Statistische Erfassungen von Kosten des Open Access je nach Lösungsweg und Methode – Grüner Weg für Zweitveröffentlichungen und direkte Veröffentlichungen auf dem Goldenen Wege – liegen der Kommission nicht vor:

*„Eine abschließende Abschätzung der Ressourcen ist aufgrund fehlender Studien aktuell nicht möglich. Für Aufbau und Betrieb von Repositorien-Infrastrukturen und korrespondierende Dienste fallen neben Anfangsinvestitionen dauerhaft Ressourcenbedarfe insbesondere in den Bereichen qualitätssicherndes Content-Management, Erschließung und Publikationsunterstützung für die Klientel an. Eine präzise Zuordnung von Ausgaben für genuine Open-Access-Aufwände ist allerdings schwierig, da die Aktivitäten – insbesondere im institutionellen Kontext – zumeist integraler Bestandteil der lokalen Informationsinfrastruktur bzw. des Publikationsmanagements und des Berichtswesens sind.“* [KII, 41]

Die weltweit sehr heterogenen Open-Access-Publikationskosten werden in Deutschland durch eine Grundförderung der Forschungsinstitutionen abgestützt, welche sich ihrerseits aus nicht unerheblichen Fördergeldern speist. Hierzu kann aus den Programmen der DFG folgender Überblick gegeben werden:

*„Seit dem Jahr 2006 hat allein die DFG über 5 Mio. EUR für Projekte zum Aufbau und zur Vernetzung von Repositorien, ca. 2 Mio. EUR für Projekte zur Information über Open Access, ca. 1,5 Mio. EUR für die Anschubfinanzierung innovativer Open-Access-Zeitschriften und im Jahr 2010 ca. 530.000 EUR für die Finanzierung von Open-Access-Publikationsgebühren an 12 Universitäten bereitgestellt. Hierzu müssen des Weiteren die vergleichsweise geringeren Investitionen in die DFG-Förderung wissenschaftlicher (Open-Access) Zeitschriften in Höhe von ca. 800.000 EUR seit 2008 sowie die Unterstützung auch von Open Access berücksichtigenden Vorhaben im Bereich von Forschungsdaten gezählt werden.“* [KII, 42]

Hinzu kommen die kaum zu beziffernden Förderpauschalen zur Deckung von Publikationskosten in Einzelprojekten, die aus einer institutionellen Grundförderung kaum zu stemmen sind [KII, 41f.].

Im Falle des Forschungsdatenmanagements sieht die *KII* vornehmlich Handlungsbedarf bei Aufbau und Dauerbetrieb von fachspezifischen Forschungsdateninfrastrukturen und Datenzentren. Einschließlich der Forschungskomponente wird für ein datendokumentierendes Zentrum mit durchschnittlich € 3,5-4 Mio. p.a. gerechnet. Bar jeder Information, welchen Umfang (Mitarbeiterzahl, Datenmenge, Dienstleistungen, Betriebskosten) ein solches Zentrum haben könnte, wird allein die Datenpflege mit einem Anteil von 5-10 % der zu erwartenden Gesamtkosten beziffert [KII, 44f.]

Ein anderer Zugang besteht in der Erarbeitung eines Kostenmodells mittels Aufgliederung des LZA-Prozesses in seine einzelnen Arbeitsschritte. Die Untersuchungen des *KoLaWiss*-Projekts zielen in eine ähnliche Richtung. Auch hier wird davon ausgegangen, dass ein beachtlicher Teil der Kosten der LZA durch eine nachhaltige Verfügbarkeit ihrer Dienste verursacht wird – eine exponentielle Korrelation von Kosten und Grad der Verfügbarkeit wird angenommen [KLW, 5]. Zu den Kostenarten werden im Wesentlichen Personal-, Investitions-, Dienstleistungs- und Infrastrukturkosten gezählt. Diese werden ebenfalls Tätigkeitsfeldern

zugeordnet – hier Acquisition, Ingest, Bit-stream Preservation, Content Preservation und Access als Arbeitsschritte der LZA [KLW, 6].

Für die Datenerhebung in Messungen, Befragungen, Editionen u.Ä. fallen nach *KoLaWiss* keine relevanten Kosten an, da der Selektionsprozess der Daten für eine Einlagerung in das Langzeitarchiv in der Regel durch den wissenschaftlichen Nutzer selbst vorgenommen wird. Es erübrigen sich demnach qualitative Prüfverfahren zur inhaltlichen Validität der Daten, wie sie zu Beginn des Speicherprozesses vorgenommen werden könnten. Dies gelte ebenso für eine (urheber-)rechtliche Überprüfung der Daten hinsichtlich Rechten Dritter, da Forschungsdaten die notwendige Schöpfungshöhe oft nicht erreichten [KLW, 7]. Trifft dies für Roh- und Messdaten sicherlich zu, sind an dieser Stelle Einschränkungen für etwa personenbezogene oder sprachwissenschaftliche Daten zu machen.

Der Ingest-Prozess, also die Einspeisung der Daten in ein Archiv, sind in der Regel Personalkosten anzurechnen – arbeiten die zuständigen Mitarbeiter hier mehrere Arbeitsschritte ab wie Gewährleistung der (technischen) Qualität, Pflege von Metadaten, Einspeisung der Daten in das Langzeitarchiv, Sicherung der Aktualität und Integrität von Dateninhalt und -format sowie Referenzier- und Auffindbarkeit [KLW, 8]. Insgesamt wird der Anteil der beiden Kostenarten Acquisition und Ingest mit ca. 42% der Personalkosten insgesamt beziffert, wobei der Ingest, die Datenerhebung nicht berücksichtigend, aufgrund seiner hohen Arbeitsintensivität den Löwenanteil einnehmen dürfte [KLW, 8].

Langfristig betrachtet gewinnen jedoch die Aufwendungen für Content Preservation und Bit-stream Preservation einen zunehmend höheren Anteil am Gesamtaufwand. Die Bit-Stream-Preservation, bestehend im Wesentlichen aus der Migration von Daten im Falle der Überführung in andere Speicherformate und der Anwendung neuer Technologien [KLW, 8], umfasst einerseits Einmalkosten für Hard- und Software sowie Kosten der Anpassung an geänderte Anforderungen. Zur Konkretisierung werden die Kosten anhand der Tarife am Göttinger Standort exemplifiziert, wo die Bit-stream Preservation durch die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)<sup>5</sup> als Dienstleister vorgenommen wird. Hierbei wird mit einem Durchschnittspreis von € 0,39 für 1 GB Daten bei der GWDG gerechnet [KLW, 8-10].

Analog zur Bit-stream Preservation stellt die Content-Preservation die langfristige Nachnutzung zur Wahrung der inhaltlichen Interpretierbarkeit der Daten sicher. Abhängig von den Anforderungen der Nutzer werden die Kosten für diesen und den Schritt der Bit-stream Preservation nach Beargie mit 23% der Gesamtkosten der LZA angenommen. [KLW, 13]

Um den Zugriff (Access) auf die Daten der Nutzer und deren Nachnutzung sicherzustellen, ist mit Aufwenden für Supportdienste zu rechnen. Obgleich eine eindeutige Trennung der dediziert für den Zugriff anfallenden Kosten vom allgemeinen Support kaum möglich ist, können im Falle der GWDG als Referenz folgende Eckdaten ersatzweise angenommen werden:

*„Zur Bearbeitung der Anfragen wird am 1st-Level Helpdesk 1 FTE einer technischen Personalkraft eingesetzt. Die Bearbeitung der Anfragen am 2nd-Level Helpdesk nimmt 0,25 FTE einer wissenschaftlichen Personalkraft und 0,5 FTE eines Technikers.“* [KLW, 14]

---

<sup>5</sup> [http://www.gwdg.de/index.php?id=wir\\_ueber\\_uns](http://www.gwdg.de/index.php?id=wir_ueber_uns)

Der Access nimmt weiterhin einen Anteil von 35% der Gesamtkosten des LZA ein [KLW, 15].

Recht analog hierzu verteilt BEUCKE bei der Ausformung eines Geschäftsmodells für die LZA die zu erwartenden Kostenarten auf die Bereiche Initiale Kosten (Projektplanung), Beschaffungskosten (Anschaffungskosten, Schaffung von Personalstellen, Neubau von Datenzentren) sowie Betriebskosten (LZA-System). Diese Kosten variieren je nach Datenaufkommen, Komplexität der Datenformate, Beschaffenheit der Metadaten, Anforderungen von Nutzerseite an die Zugriffsmöglichkeiten und rechtlichem Kontext [Beucke, 27]. Ohne eine nähere Bezifferung vorzunehmen werden diesen Bereichen die Kostenarten Creation or Purchase (abweichend zu *KoLaWiss* Kosten bei der Datengenerierung), Acquisition (Datenerfassung), Ingest, Bit-stream Preservation, Content Preservation, Access und Lifecycle Cost zugeordnet. Zu den Lebenszykluskosten eines Datenobjekt zählt BEUCKE die Aufwendungen für Qualitätskontrolle, Anreicherung mit Metadaten, Datenablage, Bestandsaktualisierung und Referenzierung [Beucke 27f.]. Darüber hinaus werden Betriebskosten identifiziert. Hierzu werden administrative, system- bzw. infrastrukturbezogene Kosten sowie Kosten der ökonomischen Anpassung genannt [Beucke, 29]. Ein letzter Kostentreiber stellen Art und Umfang der zu speichernden Forschungsdaten dar, sodass er die Standardisierung von Datenformaten empfiehlt:

*„Danach haben die Datei-Formate PDF, TXT und HTML ca. 85% aller Objekte abgedeckt, waren damit jedoch für nur 7% der Kosten verantwortlich. Währenddessen wurden 41% der Kosten durch die 12 am wenigsten verwendeten Formate (0,1%) hervorgerufen.“* [Beucke, 30]

Für das Forschungsdatenmanagement und insbesondere für die LZA haben die bisherigen Studien wenig überraschend festgestellt, dass die Personalkosten den größten Anteil innerhalb der gesamten Aufwendungen einnehmen. Nichtsdestotrotz lässt sich der Kostenaufwand durch Steuerung der Prozesse und Vorgehensweise optimieren: Kosten entstehen insbesondere dann, wenn nichtaufbereitete Rohdaten (keine Standards, keine ausreichende Metadatenausstattung) in das Langzeitarchiv aufgenommen werden sollen. In diesem Fall steigt der Arbeitsaufwand für den Daten-Ingest gegenüber der Kuration deutlich an. Unterstützt werden kann dieser Prozess somit durch eine präventive Aufbereitung des Datenbestandes und eine weitgehende Automation des Aufnahmevergangs. Was bleibt, ist dennoch die kontinuierliche Qualitätskontrolle eingehender Daten, um Fehlerhäufungen und in der Folge nachlassende Verlässlichkeit der aus den Daten generierten Ergebnisse zu vermeiden [Rad2, 31].

Zusammenfassend lassen sich die Kostenarten für LZA wie folgt aufgliedern:

*„In der Studie wird eine Verteilung der Kosten wie folgt dargestellt: 42% der entstehenden Kosten für eine Langzeitarchivierung fallen auf die Stufen Erfassung und Aufnahme der Daten, 23% fallen auf die Stufen Bitstream-Preservation und inhaltliche Langzeiterhaltung an und 35% der Kosten entstehen beim Zugriff“* [Beucke, 30].

### 3. Kostenkalkulation für TextGrid

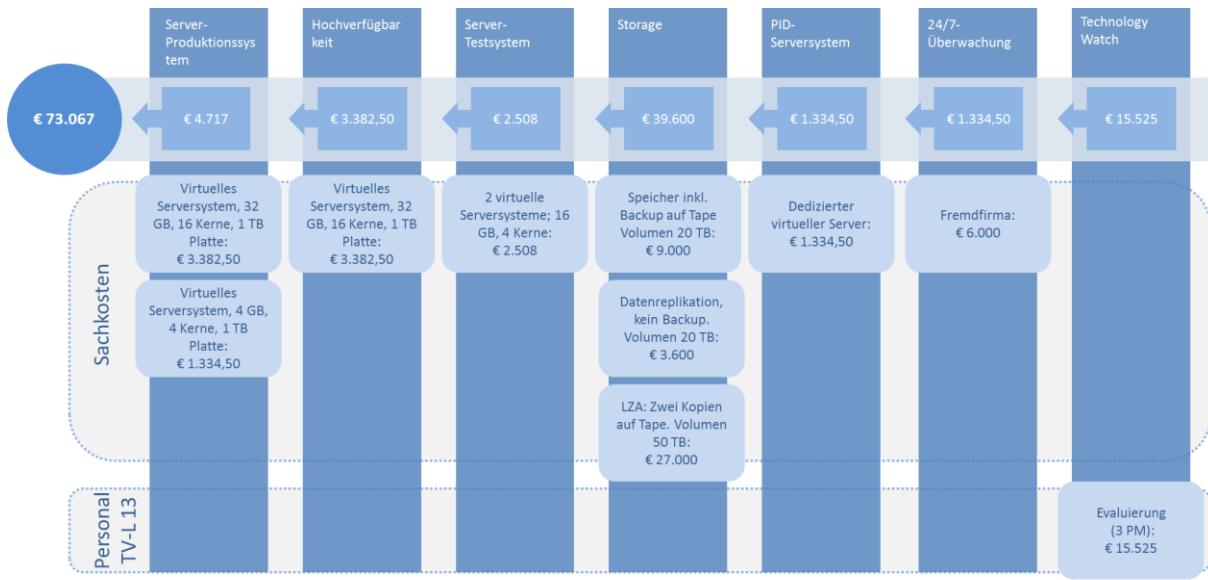
Für die Erstellung eines Kostenmodells und die daraus abgeleitete Kalkulation der Kosten sind folglich einzelne Faktoren zu identifizieren, für die Geldaufwendungen entstehen. Zwar wurde dies, wie gesehen, für den Prozess der LZA eingehend getan, für den Bereich der VFU betritt man hingegen Neuland. Grundsätzlich trifft auch bei der Erarbeitung eines Kostenmodells für eine VFU die Unterscheidung zwischen Initial-/Entwicklungskosten und Folge-/Betriebskosten zu. Für das nun für TextGrid zu entwickelnde Modell stehen jedoch die Betriebskosten einer funktionsfähigen VFU im Fokus, wurden die vorgelagerten Kosten des Aufbaus doch bereits in den projektbasierten drei Förderphasen aufgebracht. Für ein späteres Finanzierungskonzept ist demnach zunächst mit einer institutionellen Finanzierung zu rechnen. Die im Laufe des Betriebs anfallenden Anpassungen der bestehenden Architektur zur Erfüllung von Nutzerwünschen und das Aufgreifen neuester technischer Möglichkeiten, soll ein solches Modell nicht berücksichtigen. TextGrid hat sich vielmehr entschieden, solche Expansionen der VFU künftig über zusätzliche Projektmittel vorzunehmen. Eine erweiterte Funktionalität und geänderte Service Level Agreements befreien TextGrid allerdings nicht von der Notwendigkeit der Anpassung seines Kostenmodells für den Betrieb. Geänderte Anforderungen ziehen andere Kosten nach sich, sodass das vorliegende Modell lediglich den Istzustand abzubilden vermag.

Für die Identifizierung einzelner Kosten orientiert sich TextGrid an der von *Radieschen* vorgeschlagenen induktiven Methode [Rad1, 9]. Es wurden also sämtliche Dienste und Werkzeuge, die in einen Dauerbetrieb überführt werden sollen, aufgelistet und die momentan für ihre Inbetriebhaltung notwendigen, durch Monitoring der jeweils zuständigen Stellen errechneten Geldmittel ermittelt. (Weiter-)Entwicklungskosten wurden hierbei folglich nicht berücksichtigt. Die Kosten selbst untergliedern sich in Personal- (berechnet nach den von der DFG zugrunde gelegten Mittelwerten und zugehörigen Personenmonaten) und Sachmitteln (berechnet nach den im Projekt derzeit notwendigen Aufwänden).

Für die Untersuchung wurde folgende Nomenklatur aufgewandt:

- **Kostenkategorie** bezeichnet den Serviceaspekt einer VFU (Rechnerbetrieb und LZA; Tools, Archiv, Nutzerbetreuung sowie Organisation und Verwaltung)
- **Kostenkomponente** bezeichnet die in der jeweiligen Kategorie anfallenden Dienste und Tools.
- **Einzelposten** bezeichnet die zur Erhaltung der Funktionalität notwendigen Arbeitsschritte.

#### 3.1. Kostenkategorie Infrastruktur



**Abb. 1:** Kostenarten der Kostenkategorie Infrastruktur

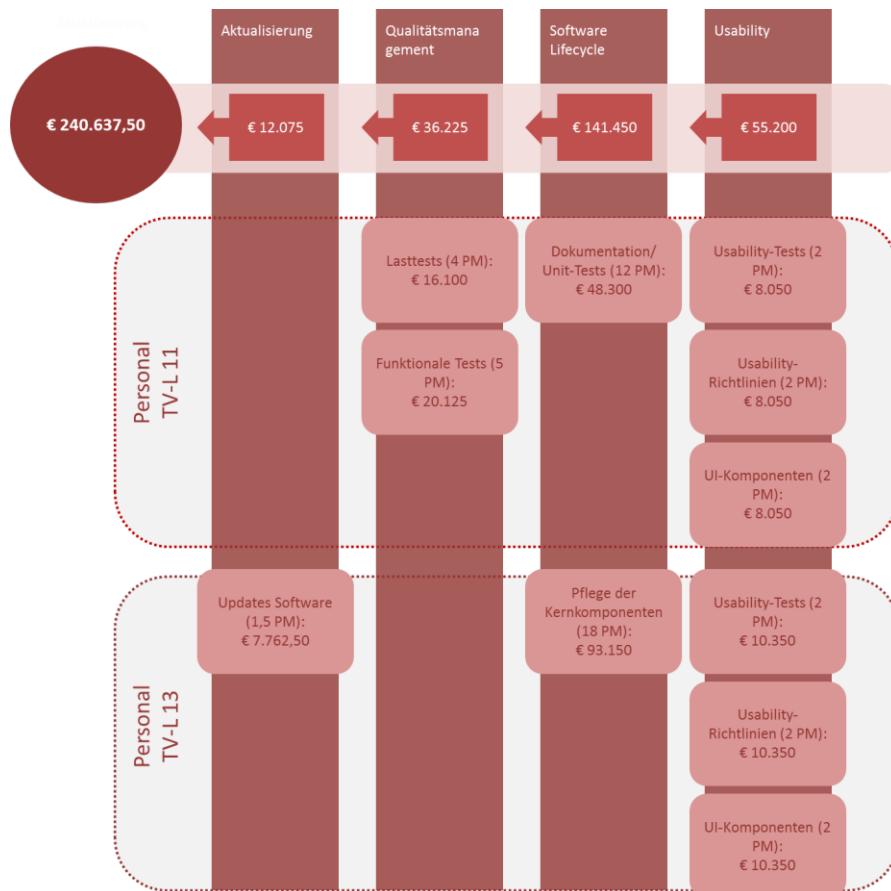
Die Kategorie „Infrastruktur“ beschreibt die technische Basis der VFU TextGrid, die im Wesentlichen auf dem Betrieb der Hardwarekomponenten und der technischen Arbeitsschritte des Datenhostings beruht. Dieser Aspekt des technischen Betriebs wird nach der momentanen Planung durch einen zentralen Ressourcenanbieter erbracht, der die einzelnen Leistungen in Vollkosten abrechnet. Vor diesem Hintergrund wurde die Kalkulation der Kosten durch den derzeitigen Vertragspartner von TextGrid, die GWDG, durchgeführt.

Unter die Kategorie „Infrastruktur“ fallen insbesondere folgende Kostenkomponenten (vgl. Abb.1):

- Betrieb eines Produktionssystems: kalkuliert wird derzeit mit Einzelposten aus zwei virtuellen Servern zu je 32 bzw. 4 GB im Goldstandard und je 1 TB an Speicherkapazität.
- Redundanz des Produktionssystems durch Betrieb eines hochverfügbaren Systems an einem zweiten Standort zu 32 GB im Goldstandard und 1 TB Speicherkapazität.
- Betrieb eines Testsystems aus zwei virtuellen Serversystemen zu 16 GB im Bronzestandard.
- Datenspeicher: enthält sowohl den Datenspeicher als solchen als auch den Speicherbedarf für die Datensicherung; je zwei Volumes zu 20 TB und das Langzeitarchivierungssystem in zwei Kopien zu je 50 TB.
- Persistente Identifikatoren (PIDs): Server zum Betrieb eines Handle-Servers für mindestens 10 Mio. PIDs zu den gespeicherten Datenobjekten.
- 24/7-Überwachung: Ständige Überwachung der Systemfunktionalität, die von einem externen Anbieter erbracht wird. Die Überwachung per se ist kostenfrei, nicht jedoch der Einsatz bei Systemausfall. Es wird mit einem Jahresaufwand von pauschal 20 Stunden bei einem Stundensatz von € 300 kalkuliert.
- Technology Watch: Zur Sicherstellung der technischen Anschlussfähigkeit der TextGrid-Architektur werden kontinuierlich technische Entwicklungen und daraus

resultierende Anforderungen überwacht und Vorschläge für mögliche Anpassungen und Implementierungen erarbeitet.

### 3.2. Kostenkategorie Betrieb/Softwarepflege Lab



**Abb. 2:** Kostenarten der Kostenkategorie TextGridLab

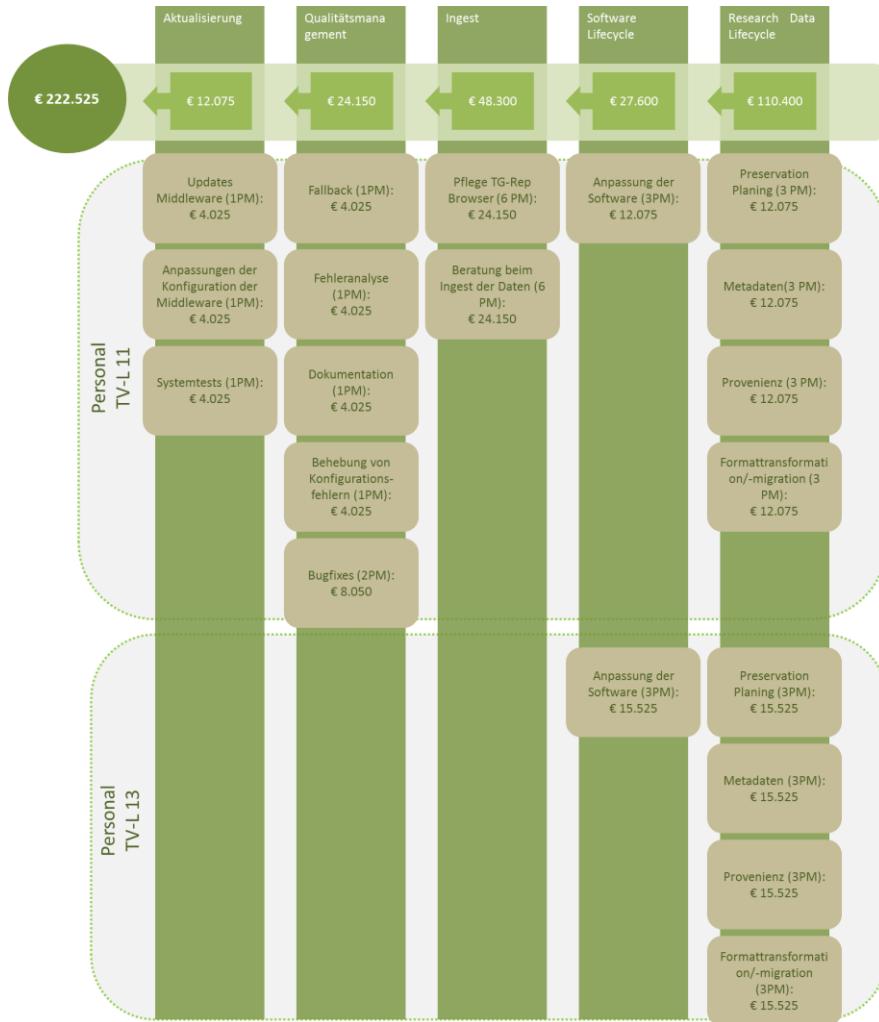
Das TextGrid Lab dient als Einstiegspunkt in die VFU und stellt Werkzeuge und Services in einer intuitiv bedienbaren Software bereit. Die Kostenkalkulation berücksichtigt lediglich die Pflege der Software nach aktuellem Stand. Kosten für eine Neu- oder Weiterentwicklung einzelner Werkzeuge sind hierbei nicht inbegriffen, sondern sollen aus zusätzlich eingeworbenen Drittmitteln finanziert werden.

Unter die Kategorie „Betrieb/Softwarepflege Lab“ fallen insbesondere folgende Einzelaufgaben (vgl. Abb. 2):

- Aktualisierung: Anpassung der vorhandenen Software nach Nutzeranforderungen und technischem Stand. Planung und Durchführung der Veröffentlichung der verbesserten Komponenten von der Konzeptions- bis hin zur Nutzungsphase. Gerechnet werden mit 2,5 PM in TV-L 13.

- Qualitätsmanagement: Erkennen und Beheben von Fehlern sowohl in Prozessen als auch in Software. Dokumentation und Publikation der Verfahren und Test einzelner Software-Komponenten auf Funktionalität sowie routinemäßig wiederholte Tests der Modifikationen der Gesamtsoftware. Gerechnet werden mit 4 PM in TV-L 11 für Lasttests und 5 PM TV-L 11 für Funktionale Tests.
- Usability: Test der Gebrauchsfähigkeit der Software für potenzielle Nutzer. Entwurf und Überarbeitung von Richtlinien zur Erkennung der Gebrauchsfähigkeit anhand nutzerdedizierter Kriterien. Konzeption und Anpassung der Bedienelemente der Software unter Anwendung der erarbeiteten Kriterien. Gerechnet werden mit 6 PM in TV-L 13 und 6 PM in TV-L 11.
- Software Lifecycle: Integration extern entwickelter Software(-komponenten) in die bestehende TextGrid-Architektur. Anpassung der bestehenden Software an veränderte Anforderungen sowie an neue Standards und Methoden. Folgende Tools werden in den Dauerbetrieb übernommen:
  - Text-Bild-Link-Editor 2: PM TV-L 13
  - Text-Text-Link-Editor: 3 PM TV-L 13
  - Metadateneditor: 1 PM TV-L 13
  - GUI, allgemeine Tools: 1 PM TV-L 13
  - Projektverwaltung / User Management: 1 PM TV-L 13
  - Projektbrowser / Navigator (GUI): 1 PM TV-L 13
  - XML-Editor: 3 PM TV-L 13
  - Recherchetool (Frontend TG-search): 1 PM TV-L 13
  - Portal: 1 PM TV-L 13
  - TG-publish: 1 PM TV-L 13
  - Wörterbuchnetz-Integration: 1 PM TV-L 13
  - Streaming Editor XSLT: 1 PM TV-L 13
  - Upload Tool (Import via TG-Lab): 1 PM TV-L 13

### 3.3. Betrieb Softwarepflege Rep



**Abb. 3:** Kostenarten der Kostenkategorie TextGridRep

Das TextGrid Repository ist ein Langzeitarchiv, im Rahmen dessen Forschungsdaten unterschiedlichsten Formats sicher gespeichert, publiziert und erschlossen werden können. Die Kostenkalkulation berücksichtigt lediglich die Pflege der momentan laufenden Komponenten, nicht aber die Kosten für eine Neu- oder Weiterentwicklung des Archivs. Maßnahmen zu dessen Expansion sollen aus zusätzlich eingeworbenen Drittmitteln finanziert werden. Es wird weiterhin davon ausgegangen, dass Daten-Ingest und Ausstattung mit Metadaten nicht Aufgabe von TextGrid sind. Vielmehr bietet TextGrid Maßnahmen an, um Fachwissenschaftler zu diesen Arbeitsschritten selbst zu befähigen. Hierbei entstehende Kosten [Beucke,30] sollten die Nutzer bei der Beantragung von Forschungsprojekten berücksichtigen, die sich Vorhaben des Datenmanagements unter Nutzung von TextGrid fördern lassen möchten.

Unter die Kategorie „Betrieb/Softwarepflege Rep“ fallen insbesondere folgende Einzelaufgaben (vgl. Abb. 3):

- Aktualisierung: Aktualisierung, Anpassung und Konfiguration der generischen Dienste zur Vermittlung zwischen Systemkomponenten und Anwendungen (Middleware). Durchführung von Tests der Funktionalität der Middleware. Gerechnet werden mit 3 PM TV-L 11.
- Qualitätsmanagement: Betrieb eines redundanten Systems zur Kompensation von Ausfällen. Erkennen und Beheben von Fehlern in Konfigurationen und Software, Dokumentation und Publikation der Verfahren und Test einzelner Komponenten auf Funktionalität. Gerechnet werden mit 3 PM TV-L 11.
- Ingest/Publikation/Präsentation: Betrieb des TextGrid-Rep-Browsers. Beratung und Unterstützung der Nutzer beim Einspeisen großer Datenmengen in das Archiv. Gerechnet werden mit 12 PM TV-L 11.
- Software Lifecycle: Integration extern entwickelter Software(-komponenten) in die bestehende TextGrid-Architektur. Anpassung der bestehenden Software an veränderte Anforderungen sowie neue Standards und Methoden. Gerechnet werden mit 3 PM TV-L 11 und 3 PM TV-L 13.
- Research Data Lifecycle: Planung des Aufbewahrungsprozesses einzelner Daten(-formate). Anlegen und Verwaltung von Beschreibungen der Daten (Metadaten). Dokumentation von Entstehung und Modifikation der Daten zur Sicherstellung ihrer Nachnutzbarkeit. Planung und Durchführung von Datentransformationen in andere Formate zur Sicherstellung langfristiger Lesbarkeit. Gerechnet werden mit 12 PM TV-L 11 und 12 PM TV-L 13.

### 3.4. Nutzerbetreuung

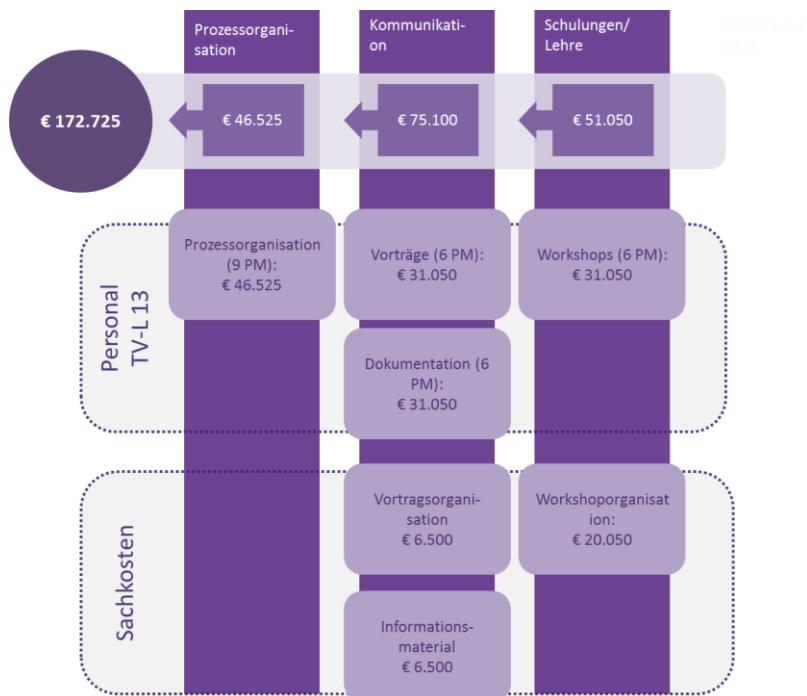


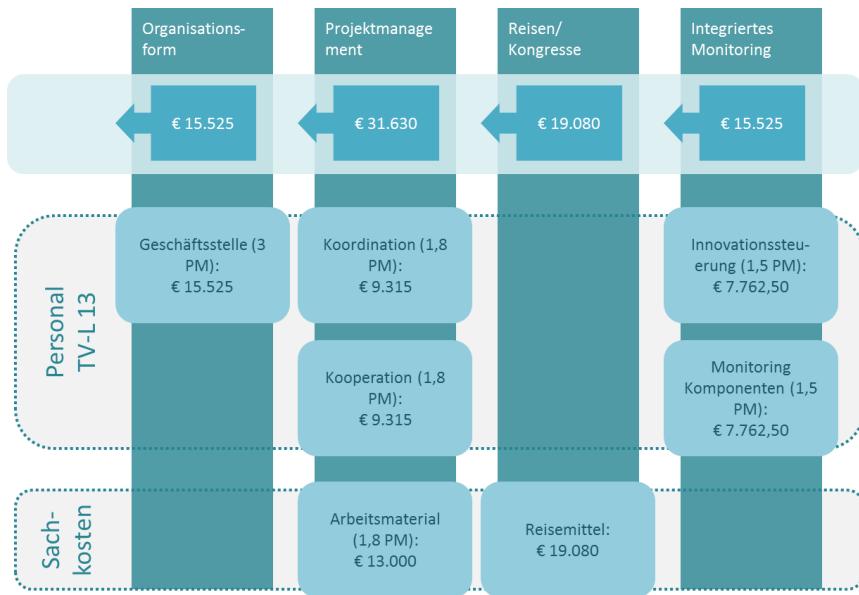
Abb. 4: Kostenarten der Kostenkategorie Nutzerbetreuung

Die Nachhaltigkeit einer VFU gründet sich nicht nur in der kontinuierlichen Pflege ihres technischen Bestandes, sondern basiert zunehmend auf ihrer Verankerung in der Forschung. Zu diesem Zweck ist TextGrid darauf angewiesen, eine breite Nutzerbasis aufzubauen, um den Mehrwert der VFU unter Wissenschaftlern zu disseminieren und ihre nutzernahe Konfiguration sicherzustellen. TextGrid verfügt daher über ein Maßnahmensortiment zur Bekanntmachung seiner Dienste, zur Unterstützung seiner Nutzer und zur Motivation neuer Nutzer.

Unter die Kategorie „Nutzerbetreuung“ fallen insbesondere folgende Einzelaufgaben (vgl. Abb. 4):

- Prozessorganisation: Prozesse im Umgang mit großen Nutzerzahlen durch Beratung heterogener Nutzergruppen, Moderation eines elektronischen Forums zum Austausch von Nutzern über die Werkzeugentwicklung sowie Pflege von Richtlinien für Qualitätsstandards für vorhandene oder zu implementierende Dienste und Werkzeuge. Gerechnet werden mit 9 PM TV-L 13.
- Kommunikation: Maßnahmen der Öffentlichkeitsarbeit bei Wissenschaftlern, Nutzern und Förderern wie Pflege der Medienpräsenz von TextGrid, Erstellung von Informationsmaterialien, Ausrichtung von Nutzer- und Entwicklertreffen, Beratung und Unterstützung sowie dauerhafte Begleitung neuer Projekte, Betreuung einer Support-List zur Betreuung einer kontinuierlichen Entwickler/Nutzer-Interaktion und Erstellung von multimedialen Tutorials zu Einführungs- und Demonstrationszwecken. Gerechnet werden mit 12 PM TV-L 13 sowie Sachmitteln für Arbeitsmaterialien.
- Schulungen/Lehre: Durchführung von abgestuften und bedarfsgerechten Schulungen zur Einführung potenzieller Nutzer in den Gebrauch der VFU. Integration der Arbeit mit VFU in Forschung und Lehre durch Modifikation von Curricula und Durchführung von Lehrveranstaltungen. Gerechnet werden mit 6 PM TV-L 13 sowie Sachmitteln für Schulungsmaterialien und notwendiger Infrastruktur.

### 3.5. Organisatorischer Betrieb



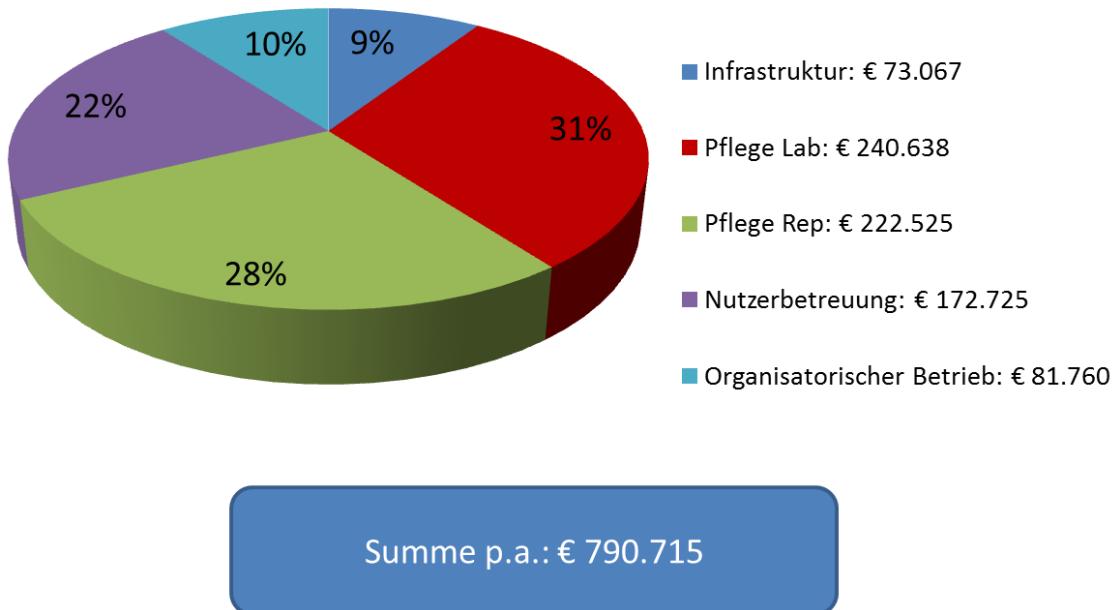
**Abb. 5:** Kostenarten der Kostenkategorie Organisatorischer Betrieb

In der Kategorie „Organisatorischer Betrieb“ sind sämtliche administrativen, organisatorischen und koordinatorischen Aktivitäten zusammengefasst.

Unter diese Kategorie fallen insbesondere folgende Einzelaufgaben (vgl. Abb. 5):

- Organisationsform: Administrative Aktivitäten rund um den Betrieb der gewählten Rechtsform. Derzeit handelt es sich hierbei um die Rechtsform des eingetragenen Vereins, im Rahmen dessen vornehmlich die Durchführung der Gremienarbeit fällt. Gerechnet werden mit 3 PM TV-L 13.
- Projektmanagement: Administrative Initiierung, Planung, Steuerung und Kontrolle der strategischen Ausrichtung der VFU. Kooperation mit relevanten Projekten, wissenschaftlichen Organisationen, Fachgesellschaften und mit Vertretern der Förderer sowie der Politik. Gerechnet werden mit 3,6 PM TV-L 13 sowie Sachmittel für Arbeitsmaterialien.
- Reisen/Kongresse: Aufwand für Dienstreisen der an der VFU beteiligten Mitarbeiter. Gerechnet werden mit € 1.600 Euro pro Jahr je Mitarbeiter für die Projekttreffen, Programmiersprints, (nationale und internationale) Konferenzen usw.
- Integriertes Monitoring: Begleitende Kontrolle der Produktqualität anhand eines speziell für die Anforderungen einer VFU entwickelten Kriterienkatalogs, Anwendung von Gouvernance-Methoden und Sondierung der Nutzerakzeptanz. Gerechnet werden mit 3 PM TV-L 13.

## 4. Fazit



**Abb. 6:** Übersicht über die Kostenarten der Kostenkategorien

Gesamtjahreskosten einer VFU waren bisher nicht bekannt. Die Verwendung der deduktiven Ermittlung der Kosten stellte sich daher als naheliegend heraus. Es wurde gezeigt, dass eine bloße Aufteilung in Initial-, Entwicklungs- und Betriebskosten vor dem Hintergrund des Werdegangs der VFU TextGrid nicht geeignet ist, um zu einer umfassenden Kostenfeststellung zu kommen. Ein zukünftiges Finanzierungskonzept sollte sich daher auf den Betrieb der VFU konzentrieren. Expansionen der Funktionalität sowie Reparatur und Neuanschaffungen sollten sich hingegen als einmalige Kosten modular an die bestehende institutionelle Finanzierung anschließen.

Analog zu den in den Studien über die Kosten der LZA gewonnenen Erkenntnissen stellen auch in TextGrid Personalkosten den Löwenanteil der finanziellen Belastungen dar. Ungefähr 85% der ermittelten Kosten entfallen auf die Gehälter der zum Betrieb zuständen Mitarbeiter. Insgesamt fallen zum Betrieb der VFU in ihrem aktuellen Stand 143 PM, also knapp 12 FTE, an. Da TextGrid jedoch ein serviceorientiertes Instrumentarium der Forschung darstellt, musste sich eine Kalkulation an den einzelnen Funktionalitäten und Angebotsbereichen der VFU ausrichten. Demnach entfallen fast ein Drittel der Gelder jeweils auf die Bereithaltung der Tools zur Erzeugung, Bearbeitung und Edition der Forschungsdaten im TextGrid Lab (31%) und die Pflege der Daten im TextGrid Rep (28%). Einen weiteren großen Anteil nehmen die Kosten für die Betreuung, Beratung und Befähigung der Nutzer ein (22%). Die übrigen 19% verteilen sich auf die in Vollkosten gerechneten Ausgaben für die Dienstleistungen eines externen Anbieters für Rechnerkapazität und Speicherplatz sowie den organisatorischen, administrativen und koordinatorischen Anteil des Betriebs (vgl. Abb. 6).

Zusammengenommen setzt TextGrid auf ein modulares Betriebskonzept mit autonomen Servicekomponenten. Um den Bedarf der wissenschaftlichen Community sinnvoll und vollständig abdecken zu können, empfiehlt sich ein auf eine gemeinsame organisatorische Basis sowie einer hinreichend ausbaufähigen, zuverlässigen aber kostengünstigen Hardwareausstattung gestützter Dauerbetrieb im Zusammenspiel aus einem Anwendungslabor für Forschungsdaten, einem Langzeitarchiv und beständiger wissenschaftlicher Begleitung des Nutzers. Diese Modularisierung entspricht der Philosophie des Projekts TextGrid hinsichtlich technischer, fachwissenschaftlicher und organisatorischer Nachhaltigkeit einer VFU. TextGrid bietet damit eine flexible und offene Struktur, die, den nachhaltigen Bestand aller Komponenten vorausgesetzt, nicht nur den Anforderungen von Nutzern aus dem Bereich der Digital Humanities an ein modernes Forschungsdatenmanagement entgegenkommt, sondern den Wissenschaftlern auch eine organisatorisch stabile Umgebung vorhält und sie mit den teils komplexen Herausforderungen im Umgang mit Forschungsdaten nicht allein lässt.

## 5. Literatur

**[Beucke]** Daniel Beucke: Geschäftsmodelle für die digitale Langzeitarchivierung. Das Beispiel Forschungsdaten, Berliner Handreichungen zur Informationswissenschaft 278 (2010); <http://edoc.hu-berlin.de/series/berliner-handreichungen/2010-278/PDF/278.pdf>

**[Beargie]** Neil Beargie, Julia Chruscz, Brian Lavoie: Keeping Research Data Safe. A cost model and guidance for UK universities (2008); <http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>

**[DFG]** Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft" - Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Ergänzte und aktualisierte Empfehlungen (2013);

[http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf)

**[GESIS]** Uwe Jensen: Leitlinien zum Management von Forschungsdaten. Sozialwissenschaftliche Umfragedaten (2012); [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2012/TechnicalReport\\_2012-07.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2012/TechnicalReport_2012-07.pdf)

**[KII]** Gesamtkonzept für die Informationsinfrastruktur in Deutschland (Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder) (2011); [http://www.allianzinitiative.de/fileadmin/user\\_upload/KII\\_Gesamtkonzept.pdf](http://www.allianzinitiative.de/fileadmin/user_upload/KII_Gesamtkonzept.pdf)

**[KLW]** Frank Dickmann: AP5 - Kosten der elektronischen Langzeitarchivierung. (2009).  
[http://kolawiss.uni-goettingen.de/projektergebnisse/AP5\\_Report.pdf](http://kolawiss.uni-goettingen.de/projektergebnisse/AP5_Report.pdf)

**[Rad1]** Torsten Rathmann: Projekt RADIESCHEN. Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur. Preise, Kosten und Domänen. Entspricht dem Report D4.3 „LZA-Kostenstruktur“ nach Projektantrag (2013); [http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:117052:2/component/escidoc:117199/ProjektRadieschen\\_Preise\\_Kosten\\_und\\_Domaenen.pdf](http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:117052:2/component/escidoc:117199/ProjektRadieschen_Preise_Kosten_und_Domaenen.pdf)

**[Rad2]** Torsten Radmann: Projekt RADIESCHEN. Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur Kostenverteilung und Risiken Entspricht dem Report D4.2 „2. Entwurf LZA-Kosten“ nach Projektantrag (2013); [http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:117203:1/component/escidoc:117202/ProjektRadieschen\\_Kostenverteilung\\_und\\_Risiken.pdf](http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:117203:1/component/escidoc:117202/ProjektRadieschen_Kostenverteilung_und_Risiken.pdf)

**[WiSta]** Markus, Zwick: Forschungsdatenzentren –Nutzen und Kosten einer informationellen Infrastruktur für Wissenschaft, Politik und Datenproduzenten, in: Wirtschaft und Statistik 12/2006 (2006), 1233-1240; [https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaDezember06.pdf?\\_\\_blob=publicationFile](https://www.destatis.de/DE/Publikationen/WirtschaftStatistik/Monatsausgaben/WistaDezember06.pdf?__blob=publicationFile)

**[WissGrid]** Frank Dickmann, Norman Fiedler, Bernadette Fritzsch, et al. Aspekte der Implementierung eines Betriebsmodells für Virtuelle Forschungsumgebungen (2011); [http://www.wissgrid.de/publikationen/deliverables/wp1/WissGrid\\_AP1\\_D15\\_v101\\_final.pdf](http://www.wissgrid.de/publikationen/deliverables/wp1/WissGrid_AP1_D15_v101_final.pdf)