Freie Universität Berlin

XML-Sicherheitsdienste für GBIF-D

Meilensteinbericht

Konzeption von Sicherheitsdiensten für die Daten- und Nachfragekommunikation in GBIF-D

Prof. Dr.-Ing. Robert Tolksdorf
Dipl.-Inf. Ekaterina Langer
Dipl.-Inf. Lutz Suhrbier

November 2004

Zusammenfassung

Dieses Dokument beinhaltet die Ergebnisse der Konzeptionsphase des ersten Arbeitspakets „Sicherheitsdienste für die Daten- und Nachfragekommunikation“ im Projekt „XML-Sicherheitsdienste für GBIF-D“. Im Rahmen der Anforderungsanalyse wurde ein Fragebogen erstellt und an alle GBIF-D-Datenanbieter zur Beantwortung verschickt. Nach Auswertung der Antworten und Analyse der Ergebnisse wurden ein Konzept zur Absicherung der Kommunikation in GBIF-D erstellt sowie relevante XML- Standards als Implementierungsbasis evaluiert.

1 EINFÜHRUNG 3

1.1 Die GBIF-D-Initiative 3

1.2 Ziele der IT-Sicherheit 4

2 ANALYSE DER BENUTZERANFORDERUNGEN 5

2.1 Erstellen des Fragenkatalogs 6

2.2 Hintergrund der Fragen und Auswertung der Antworten 7

2.2.1 Grober Projektüberblick 7

2.2.2 Dienstkontext 8

2.2.3 Akzeptanz 9

2.2.4 Datenformate und Informationsflüsse 11

3 KONZEPT ZUR ABSICHERUNG DER KOMMUNIKATION 12

3.1 Relevante XML-Sicherheitsstandards 14

3.2 Erweiterte Systemarchitektur 15

4 WEITERE ARBEITEN 17

5 LITERATURVERZEICHNIS 18

APPENDIX 19

A.1 Fragebogen 19

A.2 Ergebnisse 33

1 Einführung

„XML Sicherheitsdienste für GBIF-D“ ist ein vom Bundesministerium für Bildung und Forschung [BMBF] gefördertes Projekt mit dem Ziel das bestehende GBIF-System um Sicherheitsdienste, Rechtemanagement und Zugriffskontrolle auf XML-Basis zu erweitern. Das erste Arbeitspaket umfasst die Konzeption und die Realisierung von geeigneten Maßnahmen zur Absicherung der Daten- und Nachfragekommunikation in GBIF. Dieser Bericht beschreibt das Vorgehen und das in der ersten Phase erarbeitete Konzept für Sicherheitsdienste, die Vertraulichkeit, Integrität und Nicht-Abstreitbarkeit von Daten und Nachrichten, sowie die Authentizität der Datenquellen im GBIF-D gewährleisten. Das Dokument wurde im Rahmen des ersten Meilensteins im Projekt erstellt. Seine planmäßige Abgabe erfolgte im 11. Monat nach dem Projektstart (also im November 2004).
Kapitel 1 beginnt mit einem kurzen Überblick der Ziele der GBIF-Initiative (siehe Kapitel 1.1) sowie der schon existierenden Protokolle und Implementierungen, die zu berücksichtigen sind. Anschließend (Kapitel 1.2) werden die Grundprinzipien der IT-Sicherheit vorgestellt.
Kapitel 2 beschäftigt sich mit der Anforderungsanalyse. Die Wahl einer schriftlichen Umfrage als in diesem Fall am besten geeignete Form zur Benutzerbefragung wird begründet. Des Weiteren werden der Hintergrund der für diese Phase relevanten Fragen erläutert und die Ergebnisse analysiert.
Im 3. Kapitel wird ein Überblick XML-baiserter Sicherheitsstandards gegeben, die zur Implementierung vorgesehen wurden. Des Weiteren wird das anhand der Befragungsergebnisse erstellte Konzept vorgestellt und die erweiterte GBIF-Architektur erläutert.
Das letzte Kapitel 4 ist einem Ausblick über bevorstehende Arbeiten zur Ergänzung des Systems um Zugriffkontrolle und Rechtemanagement gewidmet.

1.1 Die GBIF-D-Initiative

Das Ziel von **G**lobal **B**iodiversity **I**nformation **F**acility [GBIF] ist die weltweite Vernetzung und Bereitstellung der vorhandenen Daten zur biologischen Vielfalt über das Internet. Forscher aus über 50 Ländern der Welt sind an der Digitalisierung konservierter Belegexemplare und der Speicherung dieser Informationen in Datenbanken, die an das GBIF-Netzwerk angebunden werden, beteiligt. Das GBIF-Internet–Portal verschafft einen einheitlichen Zugang zu diesen Biodiversitätsdaten. Um die Komplexität der weltweit verteilten Datenanbieter in der GBIF-Struktur zu reduzieren, wurden nationale Knoten gegründet. Der deutsche Knoten GBIF-D [GBIF_D] besteht aus 7 Unterknoten, die sich an größeren Organismengruppen orientieren. Ihre Aufgabe ist die Koordination der Erfassung und der Bereitstellung von in Deutschland vorhandenen biologischen Informationen.

Aus informationstechnischer Sicht stellt das GBIF-System ein modernes System zur Integration von Daten aus technisch und organisatorisch unterschiedlichen Quellen dar. In Perspektive auf eine zukünftige Nutzung des Systems erfordern die Bereiche sichere Kommunikation und Rechtemanagement weitere Arbeiten, die im Rahmen des Projektes „XML Sicherheitsdienste für GBIF-D“ durchgeführt werden. Die

geplanten Sicherheitskomponenten sollen das bestehende System erweitern und möglichst reibungslos in die vorhandene Systemarchitektur integriert werden. Aus diesem Grund wird im Folgenden ein Überblick existierender Protokolle und Softwarekomponenten gegeben, die in früheren Projekten wie z.B. BioCASE [BioCASE] entwickelt wurden.
Die BioCASE-Software erlaubt die Anbindung von beliebig strukturierten Providerdatenbanken mit Exemplardaten an das GBIF-Netzwerk. Zu den wichtigsten Komponenten zählen der PyWrapper, der UnitLoader und das BioCASE-Protokoll.
Der PyWrapper ist eine in Python implementierte XML/CGI-Datenbankschnittstelle, die einen einheitlichen Zugriff auf verschiedene Datenbanksysteme und unterschiedlich strukturierte Datenbanken ermöglicht.
Das XML-basierte Protokoll BioCASE definiert das Format der Anfragen an das PyWrapper bzw. seiner Antworten. Es abstrahiert von den zu übertragenden Daten in dem es Abfragemethoden spezifiziert, die sowohl vom Wrapper als auch vom Client verstanden werden. Auf diesem Protokoll basiert die Kommunikation zwischen der Provider-Software und den möglichen Client-Anwendungen. Aktuell sind die Abfragemethoden Search, Scan und Capabilities definiert.
Die Search-Methode bildet eine SQL SELECT-Anfrage in XML ab, wobei der Select-Teil statisch ist und alle relevanten Daten zurückliefert.
Die Scan-Methode bildet eine SELECT DISTINCT SQL-Anweisung in XML ab. Das Ergebnis ist eine Liste aller Werte für das Konzept, das durch einen XPath-Ausdruck referenziert wurde.
Mit der Capabilities-Methode kann der Client feststellen, welche Konzepte in der Provider-Datenbank abgebildet sind.
Auf der Client-Seite wird über eine grafische Benutzerschnittstelle Zugang zu den Exemplar- (Unit Level Data) und Sammlungsmetadaten (Collection Metadata) gewährt. Hauptkomponente der Client-Software ist die der UnitLoader, ein Java API, welches das BioCASE-Protokoll implementiert. Der UnitLoader generiert ein Request-Dokument und sendet es in parallelen Abfragen an alle registrierten Datenbanken.
Das BioCASE-Protokoll unterstützt die Verwendung verschiedener XML-Formate (ABCD, Darwin Core usw.) bei der Anfrage und im Ergebnis. Das Standarddatenschema in BioCASE ist das ABCD-Schema [ABCD], ein XML-basiertes Schema für primäre Biodiversitätdaten.

1.2 Ziele der IT-Sicherheit

Ziel der Integration von IT-Sicherheitskomponenten ist die Vermeidung von Schäden, die durch beabsichtigte und mit Intelligenz betriebenen Angriffe auf IT-Systeme entstehen. Sicherheit ist ein wichtiger Aspekt für alle Systeme, die online arbeiten oder Daten per Internet austauschen. Die Kommunikation ist verschiedenen Gefahren wie Abhören von vertraulichen Informationen, Verändern von Nachrichten, Fälschen von Absenderadressen usw. ausgesetzt. Um diesen Risiken zu begegnen, stellt die IT-Sicherheit folgende Kernanforderungen an ein sicheres System:

Es sollen *Vertraulichkeit* und *Integrität* der übermittelten Daten gewährleistet werden. Vertraulichkeit erlaubt Informationsgewinn nur für Berechtigte. Die Integrität stellt sicher, dass die Daten während der Übertragung nicht verändert wurden, da jegliche Datenmanipulation für den Empfänger erkennbar ist.

Das System soll die *Verbindlichkeit* von Daten und Nachrichten sicherstellen. Den Kommunikationsparteien darf es im Nachhinein nicht möglich sein, den Inhalt oder das Abschicken von Nachrichten abzustreiten.

Die *Authentizität* der Kommunikationsparteien soll gewährleistet sein, d.h. Identität und Glaubwürdigkeit der Parteien sind zweifellos feststellbar.

Mit der Erfüllung dieser Sicherheitsanforderungen wird erhofft das Bedenken bei vorsichtigen Datenanbietern zu beseitigen und ihre Bereitschaft zu erhöhen, weitere Daten an das GBIF-Netzwerk anzuschließen. Im Hinblick auf eine zukünftige Nutzung des GBIF-Systems nicht nur zur Publikation von öffentlich zugänglichen Daten ist eine Absicherung der Daten- und Nachfragekommunikation erforderlich. Das System könnte z.B. auch zur Übermittlung von Berichten an das Bundesamt für Naturschutz eingesetzt werden. In diesem Fall ist es wichtig, die Vertraulichkeit und Integrität der sensitiven Informationen zu gewährleisten, die über das Netz übertragenen werden. Des Weiteren ist die Authentizität der Datenquellen sowie die Verbindlichkeit der Berichte zu gewährleisten. Auch eine Erweiterung des GBIF-Portals, mit der Möglichkeit des modifizierenden Fernzugriffs auf Biodiversitäsdaten, ist denkbar. Für diese Zwecke benötigt das System eine Ergänzung um Rechtemanagement und Zugriffskontrolle, sowie einen sicheren Transport der Daten über das Internet.

2 Analyse der Benutzeranforderungen

Das Konzept zur Absicherung der Kommunikation in GBIF-D wurde im Anschluss an eine sorgfältig durchgeführte Anforderungsanalyse erstellt. Ziel war es, eine möglichst breite Grundlage hierfür zu schaffen. Deshalb wurden neben den GBIF-D-Projektpartnern auch Teilnehmer aus anderen ähnlichen Projekten (SYNTHESYS, BioCASE) in die Befragung einbezogen. Intensive persönliche Interviews wurden mit den Entwicklern der BioCASE-Software und des ABCD-Schemas aus der GBIF-Gruppe am Botanischen Garten und Botanischen Museum Berlin [BGBM] durchgeführt.
Wegen der großen Anzahl und räumlichen Verteilung der Projektpartner, sowie der knappen Zeit- und Finanzressourcen war es nicht möglich alle Beteiligten persönlich zu treffen. Dieses Problem wurde durch die Erstellung eines Fragebogens gelöst, der per E-Mail an alle Projektpartner geschickt wurde. Die Methode einer schriftlichen Umfrage hat den Vorteil Mittel und Zeit zu sparen. Sie erlaubt es, möglichst viele Anwender des Systems zu erreichen um deren Anforderungen und Wünsche bei der Entwicklung von Sicherheitsdiensten zu berücksichtigen. Da das Vorhaben, die Kommunikation in GBIF abzusichern, möglicherweise Bedenken bei manchen Datenanbietern bezüglich der Bereitstellung ihrer Daten auslösen könnte, waren wir in einem ausführlichen Einleitungstext und mit Erläuterungen zu den Fragen bemüht, eventuelle Zweifel aufzuklären. Dadurch erhoffen wir uns eine höhere Akzeptanz der zu integrierenden Sicherheitsdienste und des Gesamtsystems unter den GBIF-Providern. Diese ist von großer Bedeutung bei Initiativen wie GBIF, bei denen die Publikation von Daten auf freiwilliger Basis erfolgt.

Nicht zuletzt soll die Umfrage die Grundlage für einen Dialog mit den Anwendern legen, der auch in den darauf folgenden Phasen des Entwicklungsprozesses fortgesetzt werden soll.

2.1 Erstellen des Fragenkatalogs

Die Erstellung des Fragebogens hat mit Recherchen und Untersuchungen zu ähnlichen Projekten angefangen, die Web-basierte Dienste entwickeln und diese in naturwissenschaftliche Systeme integrieren.
Im Rahmen des Projekts „Framework for Europian Services in Telemedicine“ FEST [FEST] wurde ein Fragebogen [FEST_QS] entworfen, der alle relevanten Aspekte der Integration von Servicekomponenten in bestehende Umgebungen abdeckt. Dieser Fragebogen wurde als Grundlage für den Aufbau unserer Umfrage benutzt. Übernommen wurde die Strukturierung der Fragen in Gruppen, die für unsere Entwicklungsaufgaben bedeutsam sind. Zur weiteren Anpassung wurden relevante Fragen ausgewählt, die Bereiche wie Systemumgebung, Serviceentwicklung und -integration, sowie schon existierende Komponenten und Datenflüsse betreffen. Die Auswahl wurde um spezifische Fragen hinsichtlich der zu entwickelnden Sicherheitsdienste und des Biodiversitätskontextes ergänzt. Im Hinblick auf die Ergänzung des Systems durch Rechtemanagement und Zugriffskontrolle in der 2. Phase des Projekts, wurden auch Fragen bezüglich Benutzer, Rollen und Interaktionen im innerhalb des Systems miteinbezogen.
Die so zusammengestellte erste Fassung des Fragebogens wurde mit Andrea Hahn, Agnes Kirchhoff und Stephanie Sobek aus der GBIF-Gruppe am BGBM mehrmals diskutiert und überarbeitet. Dabei wurde vor allem auf die Verständlichkeit der IT-spezifischen Begriffe für nicht IT-Fachleute und die Relevanz der Fragen für den Biodiversitätskontext geachtet. Wir waren bemüht, dort wo es möglich war, Antworten zum Ankreuzen oder Auswahllisten zur Verfügung zu stellen. Zusätzlich haben wir bei allen Fragen Freitextfelder für Bemerkungen, Erklärungen und Fragen vorgesehen. Spezifische Begriffe aus den Bereichen Kryptographie und Sicherheit wurden im Glossar erläutert und dem Fragebogen beigefügt.
Die auf diese Weise überarbeitete Fassung wurde einem ausgewählten Kreis von GBIF-Projektpartnern zur Testweisebeantwortung und letzten Revision zugeschickt. Die Erkenntnisse aus den Antworten und dem Feedback dieser Testbefragung wurden in die endgültige Version des Fragebogens integriert. Diese iterative Überarbeitung des Fragenkatalogs in enger Zusammenarbeit mit Anwendern des Systems, die hauptsächlich aus dem Bereich der Biologie stammen, half uns die Eindeutigkeit der Fragen zu erhöhen und Missverständnissen bezüglich der Integration von Sicherheitsdiensten vorzubeugen.
Der komplette Fragebogen und die ausgewerteten Ergebnisse der Umfrage sind in den Anhängen A.1 Fragebogen und A.2 Ergebnisse diesem Dokument beigefügt.

Im Folgenden werden kurz die Ziele der Fragen erläutert, die für das Absichern der Kommunikation in GBIF relevant sind. Es werden die Ergebnisse diskutiert und Schlussfolgerungen gezogen.

2.2 Hintergrund der Fragen und Auswertung der Antworten

Die Befragung richtete sich an alle in GBIF-D, SYNTHESYS und BioCASE involvierten Organisationen. Es wurden per E-Mail 93 Kontaktpersonen aus 61 Organisationen angeschrieben. Damit wurde eine breite Basis für die Anforderungsanalyse geschaffen. Die meisten Befragten erhielten den Fragenkatalog in zwei aufeinander folgenden Teilen, was durch die Größe des Fragebogens begründet war.
Das Feedback war mit 61% Antwortquote bei den Institutionen für Teil 1 (siehe **Tabelle 1**) bzw. 34% für Teil 2 (siehe **Tabelle 2**) im Allgemeinen zufrieden stellend. Die beträchtlich niedrigere Antwortrate beim 2. Teil des Fragebogens ist wahrscheinlich auf die technisch ausgerichteten und zu Informatik-spezifischen Fragen zurückzuführen, die manche Teilnehmer überfordert haben. Antworten zu diesen Fragen kamen hauptsächlich von Institutionen, die eine interne IT-Abteilung haben.

2.2.1 Grober Projektüberblick

Der Fragenkatalog startet mit einer Vorabfrage (siehe Fragebogen Teil **V**), die einen groben Überblick der Projektlandschaft verschaffen soll, in die die Sicherheitsdienste zu integriert sind. Da wir neu im GBIF-Kontext sind, mussten wir in erster Linie für uns klar stellen, wer die Teilnehmer sind, welche ihre Anreize und Ziele sind (siehe Frage **V2**), sich an GBIF zu beteiligen, welche Dienstleistungen (siehe Frage **V1**) bieten sie in GBIF an. Für die Entwicklung von Sicherheitsdiensten ist es auch wesentlich, für welches Publikum die GBIF-Dienstleistungen bestimmt sind (siehe Frage **V4**).
Anhand der erhaltenen Antworten entstand folgendes Bild. Die angebotenen Dienste sind Publikation von Daten, Fernzugriff auf Daten, Einbindung von Datenquellen, Hosting von Daten, Datenanalyse und Portalentwicklung (siehe **Tabelle 3)**. Die Organisationen beteiligen sich an verschiedenen Projekten zur Online-Publikation von Biodiversitätsdaten, um von der Kombination mehrer Datenquellen zu profitieren, den Bekanntheitsgrad ihrer Sammlung zu erhöhen oder die Konsistenz ihrer Daten durch Expertenfeedback zu gewährleisten (siehe **Tabelle 4).** Aus den Antworten ist zu schließen, dass das System aktuell ausschließlich zur Veröffentlichung von allgemeinzugänglichen Daten eingesetzt wird. Dabei ist eine Absicherung der Kommunikation nicht unbedingt erforderlich.
Sehr interessant in Bezug auf die Entwicklung von Sichertheitskomponenten ist, für welche Benutzergruppen die Dienste angeboten werden. Das Spektrum der Benutzergruppen ist so breit, wie das durch das Internet erreichbare Publikum und beinhaltet von Fachleuten bis zur Öffentlichkeit, staatliche Behörden und kommerzielle Organisationen, sowohl national als auch international (siehe **Tabelle 6**).

Von besonderer Bedeutung ist die Einstellung der Projektpartner bezüglich der Integration von Sicherheitsdiensten in GBIF-D. Wird eine Erweiterung des bestehenden Systems durch Sicherheitsdienste erwünscht, welche Erwartungen sind damit verbunden und welche Anforderungen stellen die Datenprovider bzw. Anwender des Systems an die Sicherheitsdienste (siehe Frage **V6**)? Werden sich die

abgesicherten Kommunikationswege fördernd auf die Bereitschaft zur Anbindung weiterer Daten (z.B. auch sensitiver Daten) in GBIF-D (siehe Frage **V7**) auswirken?
Die Auswertung der Ergebnisse zeigt, dass über 80% der Befragten (67% der GBIF-Teilnehmer) eine Erweiterung durch Sicherheitsdienste begrüßen würden (siehe **Tabelle 7**). Der Schutz vor illegalem Datenzugriff zusammen mit Vertraulichkeit und Integrität sensitiver Daten bei der Übermittlung sind die meist gewünschten Sicherheitsmerkmale bei den GBIF-Projektpartnern. Ebenfalls von Interesse sind ein Rechtemanagementsystem sowie die Authentizität von Informationsquellen. Die Verfügbarkeit von Sicherheitsdiensten wird die Bereitschaft dem GBIF-Netzwerk weitere Daten zur Verfügung zu stellen bei 47% der Befragten (33% bei GBIF) erhöhen (siehe **Tabelle 8**). Weniger interessant für die Teilnehmer sind ein Abrechnungsdienst und die Errichtung eines Fachportals für lesende und modifizierende Fernzugriffe.

Die Gruppe **„Organisation, Rollen und Aktivitäten“** (siehe Fragebogen Teil **A**) untersucht den Prozess der Publikation von Biodiversitätsdaten, die teilnehmenden Organisationen, ihre Aktivitäten und Verantwortungsgebiete, sowie den Abhängigkeitsgrad ihrer Tätigkeiten. Die Fragen in diesem Abschnitt sind jedoch eher für die Konzeption von Rechtemanagement und Zugriffskontrolle relevant. Die Antworten finden deswegen keine detaillierte Analyse in diesem Dokument, sondern dienen lediglich zum allgemeinen Überblick der Integrationsumgebung (Prozesse, Akteure, Aktivitäten) für die Sicherheitsdienste.

2.2.2 Dienstkontext

Der Abschnitt **„Dienstkontext / Beziehungen zum Umfeld“** (siehe Fragebogen Teil **B**) untersucht rechtliche, organisatorische und sonstige Einschränkungen, die für den GBIF-Kontext typisch sind. Zu den rechtlichen Restriktionen zählen Umwelt- und Artenschutzabkommen, Datenschutz- und Telekommunikationsgesetze (siehe Frage **B1**). Wir haben uns erhofft, dass die Antworten auf diesen Fragen Hinweise zur Identifikation von sensitiven Daten und Verantwortlichkeiten geben würden. Leider ist gerade bei den GBIF-Projektpartnern mit 44% die Rate der Antwort „Ich weiß nicht. “ sehr hoch (siehe **Tabelle 9**). Die Ergebnisse verweisen primär auf Materialweitergabe-Einschränkungen (z.B. für Botanische Gärten oder mikrobiologisches Material), Arten- und Umweltschutzabkommen und Datenschutzgesetze. Die kommerzielle Benutzung der Daten wird oft mit einem Disclaimer ausgeschlossen. Die Befragten legen jedoch Wert auch auf Informationsfreiheit und Schutz der Privatsphäre. Einige geben als Kommentar den Hinweis, dass diese Aspekte von den nationalen Knoten geregelt werden müssen.

Da die Sicherheitsdienste in ein bestehendes System eingeführt werden, ist die Kenntnis der Umgebung und die Akzeptanz der Anwender von sehr großer Bedeutung. Die Vielfältigkeit der an GBIF beteiligten Institutionen und die Komplexität ihrer Strukturen machen es schwierig zu beurteilen, wie sich die Einführung von Sicherheitsdiensten auf die gängigen Arbeitsabläufe, Informationspolitiken, lokalen Standards usw. auswirken wird (siehe Frage **B2**). Die Fragen versuchen den diesbezüglichen Standpunkt der Anwender zu klären. Diese Aspekte sind wichtig für die Konzeption einer benutzerfreundlichen und leicht zu integrierenden Sicherheitserweiterung des Systems.

Die Analyse der etwas spärlichen Rückmeldung auf diese Fragen hat ergeben, dass die Einführung von Sicherheitsdiensten mit Veränderungen von gängigen Arbeitsabläufen und höherem administrativem Aufwand z. B. bei der Datenfreigabe verbunden ist. Es wird auch befürchtet, dass durch die Schilderung von möglichen Sicherheitsproblemen, die Teilnahmebereitschaft der Datenprovider sinken wird. Es wird auch darauf verwiesen, dass bei vielen Wissenschaftlern ein gewisses Misstrauen in Netzwerksicherheitsmaßnahmen existiert. So könnte die Ankündigung, auch sensitive Daten ins Netzwerk einzubinden, dazu führen, dass Datenhalter sich weigern, weitere Informationen bereitzustellen.
Diese Befürchtungen signalisieren die Notwendigkeit von Aufklärungsarbeiten in Bezug auf Fragen nach existierenden Bedrohungen und der Verlässlichkeit von Sicherheitsmaßnahmen. Als positive Folge der Sicherheitserweiterung werden der Weiterbildungseffekt und die Grundlage für weitere Mittelanforderungen gesehen.

Manche Organisationen arbeiten im Bereich IT-Sicherheit bereits mit externen Dienstleistern zusammen oder haben interne Abteilungen, die dafür zuständig sind (siehe **Tabelle 10**). Die Antworten zeigen, dass die Zusammenarbeit hauptsächlich mit instituts- oder universitätsinternen IT-Abteilungen durchgeführt wird. Diese Kooperationen und die schon eingesetzten Sicherheitsmaßnahmen (siehe Frage **B4**) müssen bei der Entwicklung der Sicherheitsdienste berücksichtigt werden. Zu den meistverbreiteten Sicherheitsmechanismen bei den Projektpartnern zählen Firewalls und Antivirus/Antispamsoftware (siehe **Tabelle 11**). Bei der Wahl der eingesetzten Software (Mail-Clients, Browser usw.) wird größtenteils auf Sicherheitsmerkmale geachtet. Sehr wenige der Befragten setzen Digitale Signaturen, E-Mail-Verschlüsselung und Zertifikate ein. Diese Ergebnisse lassen darauf schließen, dass bezüglich der Absicherung der Kommunikation wenige Maßnahmen zu berücksichtigen sind und dieser Bereich weitere Arbeiten erfordert.

2.2.3 Akzeptanz

Die unter **„Akzeptanz von GBIF-D (Vertrauen / Übereinstimmungsaspekte)“** im Teil **C** zusammengefassten Fragen diskutieren im Detail die Vorteile durch die Beteiligung am GBIF-Netzwerk für die einzelnen Organisationen (siehe Frage **C1**) und bestimmen die Notwendigkeit für die Einführung von Sicherheitsdiensten in GBIF.
Zu den direkten Vorteilen zählen effizienter, flexibler und schneller Zugriff auf biologische Daten aus zahlreichen Quellen, Vermeidung von physischem Leihverkehr sowie Austausch und Nutzung von Daten über standardisierte Formate/Schnittstellen. Dadurch ergibt sich eine breitere Basis für die Biodiversitätsforschung mit effizienteren Möglichkeiten Forschungs- und Umweltschutzmaßnahmen übergreifend zu unterstützen.
Da immer ein gewisser Zielkonflikt zwischen Sicherheitsmaßnahmen und Systemparametern wie Effizienz, einfache Bedienbarkeit usw. existiert, müssen beim Entwurf oft Kompromisse gemacht werden. Wie sicher ein Softwaresystem sein soll, hängt auch davon ab, wer es benutzt und für welche Zwecke es eingesetzt wird. In diesem Sinne ist wichtig, die Anwendergruppen zu identifizieren, deren Interessen zu berücksichtigen sind (siehe Frage **C1**). Dies sind hauptsächlich wissenschaftliche (70%) und organisatorische (50%) Interessen von Forschern und Sammlern an

Instituten, Universitäten, Museen und Ämtern. Eine starke kommerzielle Anwendung scheint bei Biodiversitätsdaten nicht verbreitet zu sein (siehe **Tabelle 12**).

Die nächste Frage (siehe Frage **C2**) ermittelt, welche Hauptanforderungen der IT-Sicherheit für die GBIF-Partner wesentlich sind, in wie weit und aus welchem Grund. Die Nichterfüllung dieser Sicherheitsanforderungen kann für die Anbieter und Benutzer eines Dienstes schwere Konsequenzen haben. Zum Beispiel führt der Verlust von Vertraulichkeit dazu, dass sensitive Daten (wie z.B. Fundorte von geschützten Pflanzen) während der Übertragung via Internet für jeden lesbar sind, der die Kommunikation abhört. Ist Integrität nicht gewährleistet, können Informationen unbemerkt, auch in bösartiger Absicht manipuliert werden.
Die Antworten (siehe **Tabelle 13**) zeigen, dass Vertraulichkeit ein wesentlicher Aspekt der Datenübertragung ist. Bei Biodiversitätsdaten überwiegen aber öffentliche Informationen, deswegen sind flexible Techniken erforderlich, die gezielt auf sensitive Daten anwendbar sind. Integrität und Verbindlichkeit der übermittelten Daten werden als wichtig für die gesamten Dokumenten und Nachrichten eingestuft. Viel Wert wird auch auf das Sicherstellen der Authentizität von Datenquellen und die Verfügbarkeit von Diensten und Daten gelegt.
Die Konsequenzen der Verletzung dieser Grundsicherheitsanforderungen (siehe Frage **C3**) reichen vom Imageverlust des Dienst- oder Datenanbieters bis hin zu finanziellen, rechtlichen oder ethischen Folgen.
Die Antworten auf diese Frage (siehe **Tabelle 14**) zeigen stark schwankende Meinungen in den verschiedenen Projekten (z.B. betrachten über 70% der SYNTHESYS-Proketpartner den Verlust der o.g. Werte als schwerwiegend). Im Gegensatz dazu sehen nur 20% der GBIF-Befragten darin eine akute Gefahr. Diese Unterschiede sind durch das Hauptziel der GBIF-Initiative begründet, Daten zur biologischen Vielfalt für alle über das Internet frei zugänglich zu machen. Natürlich handelt es sich hier nur um öffentliche Daten, deshalb wird die Verletzung ihrer Vertraulichkeit, Integrität oder Verbindlichkeit nicht als bedeutend angesehen. Die Situation ändert sich jedoch entscheidend, wenn auch sensitive Daten in das Netzwerk eingebunden oder vertrauliche Berichte ausgetauscht werden. Ein sehr drastisches Beispiel hierfür ist das Abhören des Fundortes einer geschützten Pflanzenart, das im schlimmsten Fall zum Verschwinden einer biologischen Art führen kann.

Frage **C4** betrifft die Konfiguration der Sicherheitskomponenten und die Einstellung von Sicherheitseigenschaften durch die Anwender. Unter Berücksichtigung der unterschiedlichen Sicherheitsbedürfnisse der GBIF-Projektpartner sollen die Sicherheitsdienste skalierbar entworfen werden. Die Befragten wünschen sich, das Sicherheitsniveau durch eigene Administratoren bestimmen zu können und nur die für sie sinnvollen Sicherheitsmaßnahmen einzusetzen. Dementsprechend soll die Schnittstelle zum Sicherheitsmodul über eine geeignete Visualisierung der Sicherheitseigenschaften verfügen.
Die Mittel, die zur Finanzierung der Sicherheitsdienste zur Verfügung stehen (siehe Frage **C6**), spielen eine entscheidende Rolle bei der Wahl der eingesetzten Software. Hier sind die Antworten sehr eindeutig (siehe **Tabelle 16**). Wegen der allgemeinen Knappheit der Ressourcen, gibt es wenn überhaupt sehr begrenzte Mittel für die Entwicklung und Einführung von Sicherheitskomponenten. Das bedeutet, dass für ihre Implementierung ausschließlich frei verfügbare Software und Programmbibliotheken in Frage kommen.

Die Integration von Sicherheitsdiensten macht das GBIF-System auch für andere Zwecke als die Publikation von öffentlichen Informationen einsetzbar, wie z.B. für den Austausch von sensitiven Informationen oder modifizierende Fernzugriffe auf Daten über ein Expertenportal. Diese und andere Szenarien, die eine abgesicherte Kommunikation voraussetzen, wurden in Frage **C7** angesprochen. Hier ist die Meinung der Befragten wieder mal gespalten (siehe **Tabelle 17**). Während bei GBIF nur ein sehr geringes Interesse für modifizierende Zugriffe über das Internet sowie eine Qualitätskontrolle der Daten durch Experten vorhanden ist, besteht bei SYNTHESYS ein großes Interesse an einem und der Kennzeichnung von Urheberrechten.

Für eine Evaluation des Sicherheitslevels des Systems (siehe Frage **C8**) kann es rechtliche oder verwaltungstechnische Gründe geben. Evaluationen und Zertifizierungen werden nach Europäischen Standards von staatlichen Organisationen wie z.B. dem Bundesamt für Sicherheit in der Informationstechnik (BSI) oder von kommerziellen Unternehmen durchgeführt. Die Antworten zeigen, dass 40% der Projektpartner an eine Sicherheitsevaluation interessiert sind (siehe **Tabelle 18**).

Die zwei folgenden Fragen betreffen die technischen Sicherheitsstandards (siehe **C9**), die in den einzelnen Institutionen schon eingesetzt werden. Diese sind bei der Entwicklung und darauf folgenden Einführung der Sicherheitsdienste zu berücksichtigen. Die Antworten zeigen, dass in den Projekten hauptsächlich XML-basierte Nachrichtenprotokolle wie z.B. BioCASE und DiGIR verwendet werden. Des Weiteren basieren die Datenaustauschformate ebenfalls auf XML. Diese Feststellung führt zu der logischen Wahl von XML Security Standards für die Absicherung der Kommunikation in GBIF-D. Zu den eingesetzten Sicherheitsmechanismen zählen SSL und S/Mime.

2.2.4 Datenformate und Informationsflüsse

Die letzte Gruppe **„Information“** (siehe Fragebogen Teil **D**) umfasst Fragen über eingesetzte Datenstrukturen, Informationsflüsse sowie Datenkodierungs- und Übertragungsformate in GBIF-D. Zusätzliche Aspekte stellen die Qualität der Information und sowie die Ein- und Ausgabeströme im System dar.
Ein klarer Überblick der eingesetzten Informationstypen (siehe Frage **D1**) ist wichtig für die Wahl der Standards für die Implementierung der Sicherheitsdienste. Bei den übertragenen Daten handelt es sich meistens um Text- oder Bildinformationen, sehr selten fallen auch Audiodaten an (siehe **Tabelle 19**). Nach dem Inhalt der Informationen unterscheiden die Befragten zwischen Daten über biologische Exemplare, Beobachtungsdaten, taxonomische Daten, personenbezogene Daten, Literaturdaten, GIS-Koordinaten, sowie Metadaten über Sammlungen. Also ist inhaltlich eine sehr breite Palette von Daten vorhanden, die eine jeweils unterschiedliche Schutzbedürftigkeit aufweisen. Zum Beispiel sind geografische Koordinaten von Fundorten mit hoher Wahrscheinlichkeit vertraulich zu behandeln, während bei personenbezogenen Daten über Sammler vor allem die Aspekte Privatsphäre und Datenschutz relevant sind.
Der Schutzbedarf der einzelnen Datentypen aus Anwendersicht wird in Frage **D2** angesprochen. Diese Angaben sind besonders wichtig für uns (als nicht Biologen)

um bestimmen zu können welche Informationen dauerhaft verschlüsselt werden sollten (hoher Schutzbedarf), nur bei der Übertragung verschlüsselt werden (mittlerer Schutzbedarf) oder im Klartext übertragen werden können (niedriger Schutzbedarf).
Im Allgemeinen zeigt die Rückmeldung zu dieser Frage (siehe **Tabelle 21**), dass es sich bei den meisten Daten um öffentlich zugängliche Informationen handelt. Besonders große Schutzbedürftigkeit weisen Fundorte von geschützten Arten oder Informationen aus Naturschutzgebieten auf. Auch Daten zu Sammlern bedürfen einer vorsichtigen Behandlung. Manche Befragte wünschen sich auch einen Schutz vor unberechtigtem Kopieren von Bildern.
Die folgenden zwei Fragen (siehe **D3**) betreffen die in GBIF eingesetzten Kodierungsstandards (technische wie XML, JPEG usw. und biologische wie z.B. ABCD und Darwin Core). Dieses Wissen ist wesentlich bei der Wahl von geeigneten Sicherheitsstandards für die Implementierung. Die Antworten in **Tabelle 19** zeigen, dass der meist benutzte Kodierungsstandard XML ist. Weiter werden Bildkodierungsformate wie JPEG, GIF und TIF, sowie Binär- und Audiodaten eingesetzt. Für den Datenaustausch werden XML-basierte biologische Datenschemata wie ABCD und Darwin Core verwendet. Es wird demnach vorwiegend mit XML-kodierten Daten gearbeitet, deren Besonderheiten am besten durch die Anwendung von XML-Sicherheitsstandards berücksichtigt werden können. Der Einsatz dieser Standards ist nicht nur auf XML-Daten begrenzt. Es können beliebige Datenformate verschlüsselt und signiert werden.
Es ist auch von Bedeutung in welche typischen Abläufe die Daten involviert sind (siehe Frage **D4**) und welche Datenflüsse zwischen den internen Systemkomponenten (siehe Frage **D6**) und externen Datenprovidern (siehe Frage **D7**) existieren. Die Antworten hier können grob in die Vorgänge Datengenerierung (inklusive Digitalisierung), lesender Datenzugriff (z.B. durch Suchanfragen), Datenanalyse und Bereitstellung von Daten im Netzwerk (z.B. durch den PyWrapper) unterteilt werden.

Sicherheitsdienste tragen auch zur Dienstqualität in GBIF-D bei. Frage **D5** listet einige Hauptkriterien auf, die von einem sicheren Dienst zu erfüllen sind, und fragt nach ihrer Relevanz bezüglich der Datenverarbeitung in GBIF-D. Die Befragten legen am meisten Wert auf die Gewährleistung der Datenintegrität und der Authentizität der Datenquelle bei der Übertragung (siehe **Tabelle 21)**. Um diese Anforderungen zu erfüllen, werden digitale Signaturen eingesetzt. Weiterhin sind die vertrauliche Übertragung und die Zulassung von ausschließlich autorisierten Zugriffen auf sensitive Daten wichtig. Hierfür sind selektiv anwendbare Verschlüsselungstechniken, sowie Zugriffskontrolle und Benutzer- und Rechteverwaltung notwendig. Auch die Kennzeichnung von Eigentumsrechten ist eine erwünschte Eigenschaft.
Im nächsten Kapitel stellen wir unser XML-basiertes Konzept zur Absicherung der Kommunikation im GBIF-Netzwerk vor.

3 Konzept zur Absicherung der Kommunikation

Nach einer detaillierten Analyse der Benutzeranforderungen haben wir das im Folgenden beschriebene Konzept erstellt um die Daten- und Nachrichtenkommunikation in GBIF abzusichern.

Gegenwärtig werden nur öffentliche Daten zur biologischen Vielfalt in das GBIF-Netzwerk eingebunden. Auf diese Informationen ist nur lesender Zugriff möglich z.B. über ein einheitliches Internetportal. Daten und Nachfragen werden über das Internet im Klartext ausgetauscht. Um einen vielseitigeren Einsatz dieses biologischen Informationssystems zu ermöglichen z.B. zur Anbindung kompletter Biodiversitätsdaten, inklusive sensitiver Informationen oder zur Übermittlung von vertraulichen Berichten über gefährdeten Arten, sind Maßnahmen zur Absicherung der Kommunikation notwendig. Es müssen Vertraulichkeit und Integrität der übermittelten Daten, sowie Authentizität der Datenquellen und Verbindlichkeit von ausgetauschten Nachrichten und deren Inhalten hergestellt werden. Zur Gewährleistung von Datenintegrität, Nicht-Abstreitbarkeit und Authentizität von Absendern werden digitale Signaturen eingesetzt. Die Vertraulichkeit der übermittelten Daten wird durch Verschlüsselung sichergestellt.

Die Ergebnisse der Umfrage zeigen, dass eine Erweiterung durch Sicherheitsdienste die Bereitschaft zur Publikation von Biodiversitätsdaten erhöhen wird. Die Anwender sollten aber durch gestiegene technische, administrative und organisatorische Komplexität nicht überfordert werden. Hierbei sind der unterschiedliche technische Stand der IT-Systeme und das weit variierende Niveau der IT-Kenntnisse bei den beteiligten Datenprovidern zu berücksichtigen. Um den Konfigurations- und Betriebaufwand in akzeptablen Grenzen zu halten, werden einfach zu bedienenden Administrationstools zur Verfügung gestellt. Die Installation der Sicherheitskomponenten wird in den Installationsprozess der BioCASE-Software integriert.

Die Umfrage deutet auf unterschiedlichen Schutzbedarf von Daten und variierende Sicherheitsbedürfnisse der Datengeber hin. Um diesen Tatsachen Rechnung zu tragen, haben wir uns für den Einsatz von flexiblen Sicherheitstechniken entschieden, die z.B. selektive Verschlüsselung erlauben. Weitere Schritte in diese Richtung sind die Visualisierung von Sicherheitsmerkmalen und die Möglichkeit das Sicherheitsniveau bzw. die eingesetzten Maßnahmen durch Sicherheitsadministratoren zu konfigurieren.

Ein wichtiges Designziel ist die reibungslose Integration der Sicherheitsdienste in die schon existierende BioCASE-Softwarearchitektur von GBIF. Sowohl der Datenaustausch als auch das Nachrichtenprotokoll sind XML-basiert. Aus diesem Grund haben wir uns für den Einsatz von XML-Sicherheitsstandards zur Absicherung der Kommunikation entschieden. Ein weiterer Vorteil von XML-basierten Standards ist deren Plattformunabhängigkeit und hohe Interoperabilität. Diese Eigenschaften sind sehr wichtig im Kontext von verteilten, heterogenen Informationssystemen wie GBIF. Die relevanten XML-Sicherheitsstandards werden detailliert im folgenden Unterkapitel vorgestellt.

Die Knappheit der finanziellen Ressourcen im Projekt werden wir bei unseren Entwicklungen berücksichtigen, indem für die Realisierung der Sicherheitskomponenten frei verfügbare Software und Programmbibliotheken eingesetzt werden.

3.1 Relevante XML-Sicherheitsstandards

Unser Konzept zur Absicherung der Kommunikation basiert auf folgenden XML-Sicherheitsstandards.

XML-Signature [XMLSig] ist eine Spezifikation des World Wide Web Consortiums [W3C] für die Anwendung von digitalen Signaturen im XML-Kontext. Es ist W3C Standard seit dem 12.02.2002.
Elektronische Unterschriften gewährleisten Integrität und Verbindlichkeit der übertragenen Daten. Außerdem sichern sie zweifelsfrei die Identität des Absenders.
Im Vergleich zu herkömmlichen Techniken, die nur das Signieren von kompletten Dokumenten erlauben, bietet XML-Signature den Vorteil des feinkörnigen Signierens von XML-Dokumenten. Es ist möglich Teile von Dokumenten, einzelne Elemente oder deren Inhalte zu signieren. Es ist auch möglich mehrere Dokumente in eine einzige Signatur einzubetten.
XML erlaubt große Flexibilität bei der Darstellung von Dokumenteninhalten. Zum Beispiel kann die Reihenfolge von Elementen im Dokument beliebig sein und es ist gleichgültig, ob zwischen zwei Elementen ein oder mehrere Leerzeichen vorhanden sind. Diese Flexibilität macht XML-Daten besonders schwierig zum Signieren, da bei herkömmlichen Signaturtechniken jede dieser gleichwertigen Darstellungen des Inhalts in verschiedene Fingerprints resultiert. XML-Signature löst dieses Problem, in dem jedes XML-Dokument vor dem Signieren in eine festgelegte Form überführt. Durch diese Transformation Kanonikalisierung genannt wird sichergestellt, dass XML-Instanzen vor einer Signier bzw. Verifizieroperation immer die gleiche vordefinierte Struktur haben.
Alle per URI referenzierbaren Objekte können mit XML-Signature abgezeichnet werden. Der Standard bietet drei flexible Arten die digitale Signatur mit dem Dokument mitzuliefern: e*nveloped, enveloping* and *detached Signature*. Bei den ersten zwei wird die Signatur zu einem dauerhaften Bestandteil des Dokuments, der jeder Zeit verifiziert werden kann. D*etached Signature* gibt die Möglichkeit die signierten Daten getrennt von der Signatur abzulegen.

XML-Encryption ist W3C-Standard seit dem 10.12.2002. Er spezifiziert den Verschlüsselungsprozess und die Darstellung von verschlüsselten Daten im XML-Format. Gegenüber anderen Sicherheitsstandards, die Vertraulichkeit der übermittelten Daten durch Chiffrieren gewährleisten, hat XML-Encryption den Vorteil der selektiven Verschlüsselung. Neben dem Chiffrieren kompletter XML-Dokumente erlaubt der Standard zudem die fein körnige Verschlüsselung von Elementgruppen, einzelnen Elementen und Elementinhalten. Diese Eigenschaft ist sehr nützlich im Zusammenhang von XML-Dokumenten, die hauptsächlich öffentliche Informationen und nur wenige vertrauliche Abschnitte enthalten, was bei GBIF der Fall ist.
Mit XML-Encryption ist es weiterhin möglich verschiedene Teile eines Dokuments mit unterschiedlichen Schlüsseln zu chiffrieren, so dass sie nur für bestimmte Empfänger lesbar sind. Dadurch kann bei einer Datenbearbeitung mit einigen Zwischenstationen, die unterschiedliche Berechtigungen bezüglich Teile des Dokumentes haben, Ende-zu-Ende-Sicherheit für vertrauliche Abschnitte gewährleistet werden.
XML-Encryption und XML-Signature sind in beliebiger Reihenfolge kombinierbar. Dabei bleibt die Reihenfolge der Signier- und Verschlüsselungsoperationen für den Empfänger des Dokuments transparent. Diese Eigenschaft ist hilfreich beim Umgang mit Dokumenten mit sensitivem Inhalt, die von mehreren Instanzen überprüft und

verifiziert werden. Der Vorteil für den Absender ist die Möglichkeit nur das tatsächlich gesehene (also den Klartext) zu unterzeichnen und erst dann zu verschlüsseln.
Der Apache XML Projekt stellt mit Apache-XML-Security-J version 1.1 [XML-Sec] ein frei verfügbares Java API für XML-Encryption und XML-Signature zur Verfügung. Die Programmbibliotheken stehen unter dem Apache Software License, Version 2.0 [ApLicense].

XML Key Management Spezification [XKMS] ist Kandidatstandard der W3C seit dem 5.04.2004. XKMS deckt die Verwaltung von öffentlichen und privaten Schlüsseln, die XML-Signature und XML-Encryption für ihre Ver- bzw. Entschlüsselungsoperationen benutzen, ab. Um eine Signatur zu verifizieren benötigt der Empfänger den öffentlichen Schlüssel des Absenders beziehungsweise eine Zusicherung, dass der mitgelieferte Schlüssel zum Absender gehört. Problematisch an dieser Stelle ist die große Vielfalt an Formaten und Darstellungsmöglichkeiten bei öffentlichen Schlüsseln (eingebettet in X.509 Zertifikate, identifiziert durch einen KeyName oder eine Menge von Attributen). Die verifizierende Client-Applikation muss den Umgang mit all diesen Standards und ihre spezifischen Kodierungsformate kennen, um den gewünschten Schlüssel zu extrahieren. Neben der Extraktion der öffentlichen Schlüssel ist die verifizierende Seite gezwungen Operationen wie das Validieren einer Zertifikatkette bis zu einer vertrauten Zertifizierungsstelle oder das Überprüfen von Listen mit zurückgezogenen Zertifikaten zu unterstützen. Diese oft sehr komplexen sowie zeit- und rechenintensiven Prozeduren machen den Client schwergewichtig. Um die Client-Seite zu entlasten und die Arbeit mit Schlüsseln und Zertifikaten zu vereinfachen hat die W3C die XML Key Management Spezifikation erarbeitet. Die Idee ist einen XML-basierten Trust-Service zu spezifizieren, an den sich die Client-Applikationen mit Nachfragen wenden, wenn sie einen Schlüssel oder eine Zusicherung benötigen. Die angeforderten Informationen kommen in einem vom Client gewünschten Format zurück. Um den Nachrichtenaustausch zwischen Client und Service zu spezifizieren definiert XKMS zwei XML-Protokolle: XML Key Registration Service Specification (XKRSS) und XML Key Information Service Specification (XKISS). XKRSS spezifiziert die Nachrichtenformate zur Generierung und Registrierung von öffentlichen wie privaten Schlüsseln, als auch von Vertrauenszusicherungen, die die Schlüssel mit ihren Besitzer verbinden. Anhand dieses Protokolls ist es möglich Schlüssel zurückzuziehen oder die mit ihnen verbundenen Attribute, wie z.B. Anwendungszweck oder -zeitraum zu verändern. Die zweite Spezifikation XKISS regelt das Auffinden und die Verifikation von öffentlichen Schlüsseln.
Der Einsatz eines XKMS-basierten Trust-Services zur Schlüsselverwaltung in GBIF erscheint unter Berücksichtigung der Notwendigkeit, die Client-Applikationen so einfach wie möglich zu halten, sehr sinnvoll. Allerdings ist XKMS eine relativ neue W3C-Spezifikation, für die aktuell noch keine frei verfügbare Programmbibliothek existiert. Der Apache XML Projekt plant eine Implementierung der XKMS-Spezifikation für seine XML-Security-Bibliothek.

3.2 Erweiterte Systemarchitektur

Um die Daten- und Nachfragekommunikation im GBIF-Netzwerk abzusichern, sind einige Anpassungen der vorhandenen BioCASE-Architektur erforderlich, die anhand

von **Abbildung 1** erläutert werden. Die schon vorhandenen Komponenten der BioCASE-Software sind grau gekennzeichnet.

Als erster Schritt wird eine Enc-Komponente provider- and clientseitig integriert, die Ver- bzw. Entschlüsselungsoperationen entsprechend der XML-Encryption Spezifikation durchführt. Diese Komponente ist auch für das Signieren von XML-Dokumenten gemäß XML-Signature bzw. das Verifizieren der Signaturen zuständig. Für die schon existierenden Elemente der BioCASE-Architektur sind die Verschlüsselungs-/Signieroperationen transparent. Nachdem der PyWrapper eine Nachfrage vom Client erhält, führt er die Suche in der Exemplardatenbank durch und liefert die Ergebnisse in einem dem BioCASE-Protokoll entsprechenden XML-Dokument an die Enc-Komponente weiter. Diese verschlüsselt anhand von Datenproviderpolitiken und der Identität des Nachfragers die sensitiven Abschnitte im Dokument und zeichnet es mit der digitalen Signatur des Datenproviders ab. Dadurch sind Vertraulichkeit und Integrität der Daten während der Übertragung sowie die Authentizität der Informationsquelle gewährleistet. Die so vorbereiteten XML-Dokumente werden über das Internet an den Nachfrager zurückgeschickt und erstmal von der Enc-Komponente auf der Client-Seite empfangen. Diese übernimmt die Verifikation der Signaturen und die Entschlüsselung der chiffrierten Teile. Auf dieser Weise bekommt der Client wie gewohnt eine Antwort gemäß dem BioCASE-Protokoll und die Absicherung der Kommunikation bleibt sowohl für ihn als auch für die Provider-Software transparent.

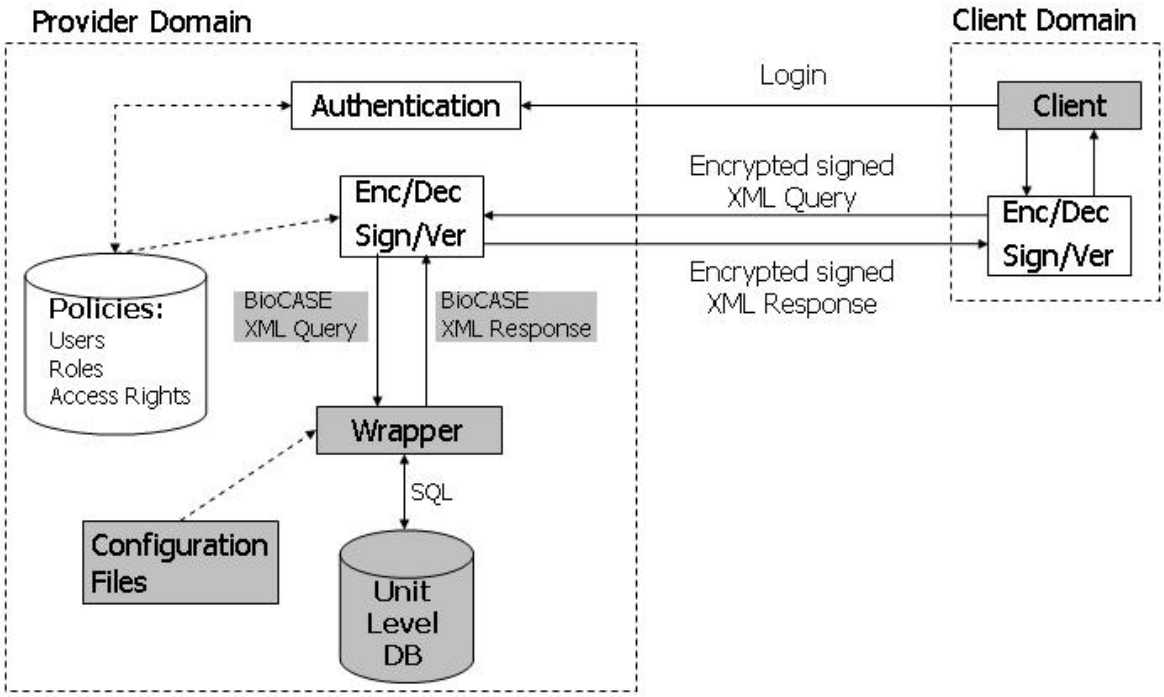

Abbildung 1 Erweiterte Systemarchitektur

Um die Identität der Anwender festzustellen, wird eine Authentifizierungskomponente integriert. Clients (Benutzer oder Programme) müssen sich authentifizieren bevor sie eine Anfrage an den Datenprovider schicken dürfen. Um den Authentifizierungsprozess flexibel zu gestalten, wird das System sowohl passwort-

als auch PKI-basiertes Einloggen anbieten. Informationen über Benutzer, Rollen, Zugriffsrechte und Berechtigungsnachweise werden von einer Rechtmanagement-Komponente verwaltet.

Des Weiteren sind einige Erweiterungen des BioCASE-Protokolls notwendig. Die erste betrifft die Abfrage der unterstützten Sicherheitsmechanismen. Für diesen Zweck wird die Methode Capabilities um einen Platzhalter für die Abfrage sicherheitsrelevanter Merkmale erweitert. Diese Anpassung erlaubt es auch solche Clients bzw. Provider an das GBIF-Netzwerk anzubinden, die über keinerlei Sicherheitsmechanismen verfügen.
Weiter ist eine Erweiterung des BioCASE-Anfrageformats um einen Login-Identifikator vorgesehen. Dieser wird dem Client nach einem erfolgreichen Authentifizierungsprozess zugeteilt und seinen nachfolgenden Anfragen beigefügt. Dadurch wird eine erneute Anmeldung des Clients bei wiederholten Nachfragen vermieden.

4 Weitere Arbeiten

Das hier vorgestellte Konzept zur Absicherung der Kommunikation in GBIF benötigt neben der Authentifizierung der Benutzer auch Zugriffskontrolle auf Ressourcen, sowie ein geeignetes Rechtemanagementsystem zur Definition von Zugriffspolitiken und der Verwaltung von Benutzern. Auch die Ergebnisse der durchgeführten Umfrage im Rahmen der Anforderungsanalyse zeigen starkes Interesse an diese Features. Aus diesen Gründen ist in der zweiten Phase des Projekts die Ergänzung der GBIF-Infrastruktur um Rechtemanagement und Zugriffskotrolle geplant. Für ihre Realisierung sind die XML-basierten Standrads eXtensible Access Controll Markup Language [XACML] und Security Assertion Markup Language [SAML] vorgesehen. XACML bietet ein XML-Vokabular zur Definition von Zugriffsrechten auf Ressourcen und ein Protokoll zur Ermittlung von Zugriffsentscheidungen. Auf dessen Basis sollen differenzierte Zugangspolitiken für Datennachfrager spezifiziert und eingesetzt werden. SAML definiert XML-basierte Nachrichtenformate für den Austausch von Authentifizierungs- und Autorisierungsinformationen zwischen Domänen in einem verteilten System. Auf dieser Weise ist es Clients mit den entsprechenden Berechtigungen möglich, nach erfolgreicher Anmeldung bei einer Domäne auf Ressourcen weiterer Domänen zuzugreifen, ohne sich erneut authentifizieren zu müssen. Basierend auf diesen beiden XML-Standards wird eine flexible Struktur von Zugriffspolitiken entworfen, bei der jeder Provider die Zugriffsrechte auf seine Daten selbst definieren kann. Nationale oder projektspezifische Politiken, deren Gültigkeit über mehrere Datenprovider reicht, werden einmalig vom nationalen bzw. Projektknoten spezifiziert und können dann von jedem Datenprovider übernommen werden.
Die Erweiterung des GBIF-Systems um Sicherheitsdienste, Rechtemanagement und Zugriffskontrolle trägt direkt zur besseren Qualität der angebotenen Informationsdienste bei und eröffnet zusätzliche Nutzungsmöglichkeiten. Die eingesetzten Sicherheitsmechanismen werden die Attraktivität der Teilnahme an GBIF für Datenanbieter, Sammlungsbesitzer und Forschungseinrichtungen erhöhen.

5 Literaturverzeichnis

[BMBF]	Bundesministerium für Bildung und Forschung: http://www.bmbf.de
[FEST]	Das FEST-Projekt: http://www.cee.hw.ac.uk/Databases/lachs/fest.html
[FEST_QS]	Der FEST Fragebogen: ftp://ftp.cs.tu-berlin.de/pub/local/kit/documents/KIT-Reports/r123/FG-AA
[BioCASE]	Das BioCASE Projekt: http://www.biocase.org/default.shtml
[GBIF_D]	GBIF-Deutschland: http://www.gbif.de/
[GBIF]	GBIF: http://www.gbif.org/
[ABCD]	Das ABCD-Schema: http://www.bgbm.org/TDWG/CODATA/Schema/default.htm
[BGBM]	Botanischer Garten und Botanisches Museum Berlin-Dahlem: http://www.bgbm.org/
[W3C]	W3C: World Wide Web Consortium (W3C); http://www.w3.org
[XMLSig]	W3C: M. Bartel, J. Boyer, B. Fox, B. LaMacchia, E. Simon: XML-Signature Syntax and Processing, http://www.w3.org/TR/xmldsig-core/
[XMLEnc]	W3C: T. Imamura, B. Dillaway, E. Simon: XML-Encryption Syntax and Processing (2002), http://www.w3.org/TR/xmlenc-core/
[XKMS]	W3C: Phillip Hallam-Baker XML Key Management Specification XKMS 2.0 (2004), http://www.w3.org/TR/2004/CR-xkms2-20040405/
[XML-Sec]	Apache XML Security, Java API: http://xml.apache.org/security/Java/index.html
[ApLicense]	The Apache Software License, Version 2.0: http://xml.apache.org/security/license.html
[XACML]	OASIS: eXtensible Access Control Markup Language TC; http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml
[SAML]	OASIS: OASIS Security Services TC: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security

Appendix

A.1 Fragebogen

Einleitung

Dieses Dokument beinhaltet einen Fragebogen, der zur Anforderungsanalyse bzgl. der Integration von Sicherheitsdiensten und Rechtemanagement in GBIF-D oder vergleichbaren Projekten (z.B. BioCASE, SYNTHESYS) entwickelt wurde. Weitere Informationen zur Integration von Sicherheitsdiensten finden Sie sind unter http://nbi.inf.fu-berlin.de/research/GBIF-D/.

Die Fragen basieren auf einem Fragenkatalog, der im Rahmen des FEST-Projekts der Technischen Universität Berlin zur Analyse eines telemedizinischen Dienstes erstellt wurde. Der Katalog wurde an die Bedürfnisse von GBIF-D angepasst.

Sie haben mit der zügigen und sorgfältigen Beantwortung des Fragenkatalogs die Möglichkeit aktiv Ihre Wünsche, Erwartungen und Anregungen bzgl. der Erweiterung von GBIF-D um Sicherheitsdienste in den Anforderungskatalog einzubringen. Gerne können Sie die Fragen auch an Ihre entsprechenden Fachabteilungen weiterleiten. Wir möchten Sie jedoch bitten, uns die beantworteten Fragebögen spätestens bis zum **24 Mai 2004** zurückzusenden, damit wir umgehend mit der Analyse beginnen können.

Der Fragenkatalog richtet sich an GBIF- und BioCASE-Datenprovider und wird an möglichst viele beteiligte Institutionen verschickt. Damit soll die Anforderungsanalyse auf eine möglichst breite Basis gestellt werden. Gegebenenfalls werden wir uns mit weiteren Fragen an Sie wenden. Diese Vorgehensweise ermöglicht es uns, dynamisch auf die erhaltenen Antworten der befragten Teilnehmer einzugehen und gezielte Nachfragen zu formulieren. Auf diese Weise möchten wir versuchen den Dialog bzgl. der Sicherheitsdienste mit allen Beteiligten über die gesamte Laufzeit des Projekts beizubehalten.

Einige Grundbegriffe der IT-Sicherheit werden im Glossar erläutert. Bitte wenden Sie sich bei auftretenden Unklarheiten oder Problemen an uns, damit wir diese umgehend beseitigen und den Fragenkatalog gegebenenfalls überarbeiten können.

Bei der Beantwortung der Fragen sind grundsätzlich Mehrfachnennungen möglich. Wenn dies nicht der Fall ist, wird explizit drauf verwiesen. Bei den Ankreuzfragen ist oft die Möglichkeit auch zur Freitextantworten gegeben. Wir möchten Sie bitten, diese Freitextfelder jedes Mal zu nutzen, wenn eine Ihrer Antworten nicht in der Ankreuzauswahl zu finden ist, Sie wichtige Details betonen bzw. verdeutlichen möchten oder aber Schwierigkeiten bei der Beantwortung auftreten.

Das Ausfüllen des Fragebogens sollte im geschützten Modus (Benutzereingabe ist nur in den grauen Feldern erlaubt) erfolgen. Dieser Modus sollte sich automatisch beim Öffnen des Word-Dokuments einstellen. Ist das nicht der Fall, dann einfach das Schlosszeichen auf der Symbolleiste oder den Menüpunkt ‚Extras/Dokument schützen' anwählen. In den grauen Textfeldern kann beliebig langer Text eingegeben werden (einfach mit der Maus drauf klicken und tippen, die Taste ‚Enter' erzeugt eine

neue Zeile). Außer den Text-Feldern werden in diesem Formular Kontrollkästchen (ankreuzen durch Mausklick) verwendet.

Wir bedanken uns im Voraus für Ihre Mühe.

Ekaterina Langer, atanasso@inf.fu-berlin.de
Lutz Suhrbier, suhrbier@inf.fu-berlin.de

Glossar:

Authentizität
Die Identität von Parteien ist durch Zertifikate in Verbindung mit Authentisierungsprotokollen zweifelsfrei feststellbar.

Autorisierung
Der Zugriff auf Daten ist nur berechtigten Benutzern erlaubt. Eine Zugriffskontrollkomponente überwacht den autorisierten Zugriff auf (Teile von) Dokumente und bietet eine Schnittstelle zum Rechtemanagement.

Digitale Signatur
Ein Dokument wird signiert indem ein für dieses Dokument eindeutiger Hashwert (digitaler Fingerabdruck) erzeugt und mit dem privaten Schlüssel des Absenders verschlüsselt wird. Dieser Hashwert wird vom Empfänger mit dem öffentlichen Schlüssel des Absenders entschlüsselt und mit dem über das empfangene Dokument erstellten Hashwert verglichen. Bei Übereinstimmung ist sowohl die Authentizität des Absenders als auch der manipulationsfreie Empfang des Dokuments gewährleistet.

Digitale Zertifikate
Digitale Zertifikate werden von vertrauenswürdigen Zertifizierungsstellen ausgestellt um die Zugehörigkeit eines öffentlichen Schlüssels zu einer Person/Organisation/Dienst zu bestätigen.

Integrität
Daten können nicht unbemerkt geändert werden. Dies wird durch die Verwendung Digitaler Signaturen sichergestellt.

Rechtemanagement
Zuordnung von Zugriffsrechten auf Objekte (Dokumente, Ressourcen) anhand von Personen-, Gruppen- oder Systemmerkmalen. Aufgrund dieser Zuordnung entscheidet die Zugriffskontrolle wer, welche Operation (z.B. Lesen, Schreiben usw.) auf welchen Objekten ausführen darf.

Schlüsselpaar
Wird in asymmetrischen Verschlüsselungsalgorithmen und beim Digitalen Signieren verwendet. Ein Schlüsselpaar besteht aus einem privaten und einem öffentlichen Schlüssel. Der private Schlüssel ist nur seinem Eigentümer bekannt. Der öffentliche Schlüssel ist allgemein bekannt (z.B. ist in öffentlichen Verzeichnisdiensten zugänglich).

Anwendung beim Verschlüsseln:
Möchte man sicherstellen, dass eine Nachricht nur von Ihrem Empfänger gelesen wird, verschlüsselt man diese mit dem öffentlichen Schlüssel des Empfängers. Nur dieser ist dann in der Lage die Nachricht mit seinem privaten Schlüssel zu entschlüsseln.
Anwendung bei Digitaler Signatur:
Der Hashwert des Dokumentes wird vom Unterzeichner mit seinem privaten Schlüssel verschlüsselt. Damit ist die eindeutige Zuordnung Dokument / Unterzeichner sichergestellt. Der Empfänger verifiziert den Ursprung des Dokuments mit Hilfe des öffentlichen Schlüssels des Unterzeichners (siehe Digitale Signatur).

Verbindlichkeit / Nicht-Abstreitbarkeit
Parteien können Nachrichten und Inhalte nicht abstreiten. Dies wird durch Digitale Signaturen in Verbindung mit Zertifikaten gewährleistet.

Verfügbarkeit
Dienste stehen permanent zur Verfügung

Vertraulichkeit
Daten können von Unbefugten nicht gelesen werden. Dies wird durch Verschlüsselung der Daten gewährleistet.

Unklare Begriffe:

Sind Sie bei der Beantwortung des Fragebogens auf unklare Begriffe gestoßen? Bitte tragen Sie diese in das untenstehende Textfeld ein. Wir werden diese Begriffe nach Möglichkeit in das Glossar aufnehmen.

Unklare Begriffe:

Name der Institution	_____
Kontaktperson	_____
E-Mail	_____
Datum	_____

Vorabfrage

Hilfe
Die Vorabfragen zielen darauf ab Informationen bzgl. der Einsatzumgebung des zu entwickelnden Dienstes zu ermitteln.

V1. Welche Art von Dienstleistungen bieten Sie innerhalb von GBIF-D an? Bitte beschreiben Sie kurz den Inhalt der angebotenen Dienstleistungen.

☐	Publikation von Daten
☐	Zugangsmechanismen
☐	Einbindung von Datenquellen
☐	Analyse von Daten
☐	Portalentwicklung
☐	Schnittstellen-Software für die Anbindung von Datenbanken
☐	sonstige Dienstleistungen:
☐	Keine

V2. Welche Ziele möchten Sie mit der Teilnahme an GBIF-D erreichen? Bitte begründen Sie Ihre Auswahl.

☐	Sammlungsdaten online verfügbar machen
☐	Erhöhung des Bekanntheitsgrades Ihrer Datensammlung
☐	Korrektheit der Daten (Experten-Feedback)
☐	Mehrwert durch Kombinationen mehrerer Datenquellen
☐	„Werbung“ für Sammlungsdaten als wichtiger Aspekt der Biodiversitätsforschung
☐	Automatisierte „Massenannotation“ als Zukunftsziel (z. B. durch maschinellen Abgleich von Koordinaten-Daten mit Ländernamen)
☐	Sonstige Ziele:

V3. Wer leistet die Finanzierung für die Entwicklung/Betrieb der(s) Dienste(s)?

	☐	Eigene Mittel
	☐	Nationale Organisation(en)
	☐	Internationale Organisation(en)
	☐	Kommerzielle Organisation(en)
	☐	Förderprogramme: ☐ BIOLOG ☐ EU-Programme (BioCASE, ENBI) ☐ GBIF-Programme ☐
	☐	Sonstige:

V4. Für welches Publikum soll der Dienst angeboten werden?		
	☐	Geschlossene Benutzergruppen bei sensiblen Daten
	☐	Fachpublikum
	☐	Öffentliches Publikum
	☐	Kommerzielle Organisationen
	☐	Internationales Publikum
	☐	Nationales Publikum
	☐	Sonstiges
V5. Seit bzw. ab wann nehmen Sie an GBIF-D teil?		

V6. Wünschen Sie sich eine Erweiterung des GBIF-Systems um Sicherheits-dienste? Wenn ja was versprechen Sie sich davon? Bitte erläutern Sie nach Möglichkeit die Hintergründe.		
	☐	ja
	☐	nein
	☐	Schutz vor unberechtigten Datenzugriff (Zugriffskontrolle)
	☐	Schutz vor Datenmanipulation
	☐	Vertraulichkeit übermittelter Daten
	☐	Möglichkeit des zielgruppen-orientierten Angebot von sensiblen Daten
	☐	Abrechnungsdienst als Teil eines möglichen Zukunftsgeschäftsmodells
	☐	Vertrauenswürdige Datenquellen
	☐	Rechtemanagement
	☐	Einrichtung eines Fachportals für einen geschlossenen Nutzerkreis (z.B. nach dem Muster des Australian Virtual Herbarium), über das Daten in größerem Detailgrad ausgetauscht werden können
	☐	Sonstiges:

V7. Würde die Verfügbarkeit von Sicherheitsdiensten die Bereitschaft erhöhen weitere Daten in GBIF-D zu veröffentlichen?
V8. Gibt es sonstige (z.B. rechtliche, wirtschaftliche) Gründe, die zukünftig die Integration von Sicherheitsdiensten in GBIF-D erforderlich machen könnten? Wenn ja, welche?

A. Organisationen, Rollen und Aktivitäten

Hilfe Dieser Abschnitt ermittelt die an GBIF-D beteiligten Akteure (Organisationen, Personen usw.), ihre Aktivitäten, Verantwortlichkeiten und der Grad ihrer Kooperation.

A1. Bitte beschreiben Sie den Ablauf der Bereitstellung von Biodiversitätsdaten ins Internet (von der Digitalisierung der Daten bis zur Veröffentlichung ins Internet).

A2. Aktivitäten / Aufgaben / Tätigkeiten
a) Für welche Aufgaben im Rahmen von GBIF-D ist Ihre Institution verantwortlich?
b) Mit welchen Institutionen arbeiten Sie zur Realisierung dieser Aufgaben zusammen?

A3. Akteure in GBIF-D
a) Wer ist in Ihrer Institution für die Bereitstellung der GBIF-D-Dienste zuständig (z.B. Rechenzentrum, Web-, Server-, Netzwerk-, Systemadministrator, Software-Entwickler, techn. / wiss. Mitarbeiter, stud. Hilfskräfte als Rolle Eingabe, Bereitstellung, Zusammenführung, Sicherung von Daten, Qualitätskontrolle, Pflege des Webangebots, Hosting von Datenbanken, Schnittstellenentwicklung als Beispiel für Aufgaben)?

Rolle (Berufsbezeichnung)	**Aufgaben**

b) Für welche Zielgruppen bieten Sie Daten oder Dienste innerhalb von GBIF-D an oder möchten Sie in Zukunft anbieten? (z.B. Fachbehörden, Ministerien, Verbände, Universitäten, Wissenschaftler, interessierte Öffentlichkeit usw.)

Rolle / Zielgruppenbezeichnung	**Aufgaben bzw. wofür werden die Daten benötigt**

A4. Welche Organisationen sind direkt an der Realisierung der von Ihnen angebotenen GBIF-D-Dienste beteiligt? Bitte beschreiben Sie deren Aufgaben bzgl. Ihrer Institution. Mit Organisationen sind unter anderem Museen, Herbarien, Institute, Ämter, Universitäten, Umweltschützer, Internetprovider, kommerzielle Organisationen gemeint.

Organisationsname	**Aufgaben**
z.B. Botanisches Museum Berlin	Host, Veröffentlichung einer Kopie der DB

B. Dienstkontext / Beziehungen zum Umfeld

Hilfe
Um Sicherheitsdienste in bestehende Strukturen des GBIF-D-Dienstes einführen zu können, ist eine Strategie zu entwickeln, die den Erfolg dieser Einführung sicherstellt. Insbesondere müssen die Sicherheitsdienste in die vorgegebene Umgebung bestehender GBIF-D-Dienste eingepasst werden.

B1. Welche rechtlichen Einschränkungen bestehen für GBIF-D?
Bitte schildern Sie kurz die Konsequenzen dieser Einschränkungen.

☐	Datenschutz- und Telekommunikationsgesetze
☐	Umwelt- und Artenschutzabkommen
☐	Materialweitergabe-Einschränkungen für botanische Gärten
☐	Haftung und Normen
☐	Informationsfreiheit
☐	Schutz der Privatsphäre
☐	Verantwortung von Einzelpersonen und Organisationen
☐	Signaturgesetz
☐	Ausschließen der Kommerzialisierung der bereitgestellten Daten z.B. durch die GNU-Copyleft-Lizenz
☐	Disclaimer für die publizierten Daten. Kann auf folgende Adresse gefunden werden:
☐	Sonstige:
☐	Weiß nicht

B2. Organisatorische Konsequenzen von GBIF-D

a) Welche positiven oder negativen Auswirkungen auf bestehende Organisationsstrukturen erwarten Sie durch die Einführung von Sicherheitsdiensten in GBIF-D an Ihrem Institut?
z.B. Auswirkungen auf:
- die bestehende Arbeitsweise, Verantwortlichkeiten
- lokale Standards und Normen
- Informationspolitik und Schulungen
- bestehende Investitionen
- weitere Mittelanforderung

b) Welche organisatorischen oder technischen Features machen evtl. Anpassungen im Sicherheitsmanagement erforderlich?

z.B. Änderung von Vorschriften oder Gesetzen, lokale Standards und Normen, Einführung von Feedback-/Annotationsmechanismen, bestehende Investitionen, (Netzwerk-)Infrastruktur

B3. Arbeiten Sie im Bereich IT-Sicherheitsmanagement bereits mit internen Fachabteilungen Ihrer Institution oder mit externen Organisationen / Anbietern zusammen?

☐	ja
☐	nein

Wenn ja mit welchen?
Bitte erläutern Sie kurz die Art der Zusammenarbeit.

	Organisation
☐	Interne Abteilung
☐	staatliche / regionale / öffentliche Institutionen
☐	Zertifizierungsstellen
☐	private Unternehmen
☐	Sonstige:

B4. Mit welchen Maßnahmen ist die Zusammenarbeit mit diesen Organisationen / Anbietern verbunden?
z.B. Firewall, verschlüsselte E-Mails, digitale Signatur, Chipkarten, Sicherheit der lokalen Rechner, Einschränkungen der eingesetzten Software

C. Akzeptanz von GBIF-D (Vertrauen / Übereinstimmungsaspekte)

C1. Vorteile / Interessen

a) Welche Vorteile sehen Sie in der Einführung / Nutzung von GBIF-D-Diensten in Ihrer Organisation?

b) Wessen Interessen sind dabei möglicherweise zu berücksichtigen?
Bitte ergänzen Sie die Antworten nach den Gegebenheiten Ihrer Institution und erläutern Sie kurz bestehende oder zukünftige Interessen.

Interessenarten	**Interessenten**
☐ Wissenschaftliche	☐ Institute ☐ Universitäten ☐ Biologen, ☐ Sammler

	☐ Sonstige:
☐ Technologische	☐ Naturwissenschaftler ☐ Forscher ☐ Umweltschutzorganisationen ☐ Sonstige:
☐ Organisatorische	☐ Museums-/ Institutleitungen ☐ Ämter ☐ Sonstige:
☐ Wirtschaftliche	☐ Unternehmen ☐ Sonstige:
Sonstige:	

C2. Welche Bedeutung messen Sie folgenden Grundwerten der IT-Sicherheit für die von Ihrer Organisation angebotenen oder genutzten GBIF-D-Dienste bei? Bitte erläutern Sie kurz Ihre Antworten und geben Sie Beispiele für sensible Daten, falls Sie die Antwort „nur für bestimmte Daten“ ausgewählt haben.

Grundwert	**Bedeutung**
Vertraulichkeit (Daten können von Unbefugten nicht gelesen werden.)	☐ unwichtig ☐ nur für bestimmte Daten z.B.: ☐ generell wichtig
Integrität (Daten können nicht unbemerkt verändert werden.)	☐ unwichtig ☐ nur für bestimmte Daten z.B.: ☐ generell wichtig
Verbindlichkeit / Nicht-Abstreitbarkeit (Parteien können Nachrichten und Inhalte nicht abstreiten)	☐ unwichtig ☐ nur für bestimmte Daten z.B.: ☐ generell wichtig
Authentizität (Identität von Parteien zweifelsfrei feststellbar)	☐ unwichtig ☐ nur für bestimmte Daten z.B.: ☐ generell wichtig
Verfügbarkeit (Dienste stehen permanent zur Verfügung)	☐ unwichtig ☐ nur für bestimmte Daten z.B.: ☐ generell wichtig

C3. Welche Folgen/Kosten hätte der Verlust der in C2 genannten Grundwerte für die von Ihrer Organisation angebotenen oder genutzten GBIF-D-Dienste?

Grundwert	**Folgen/Kosten**
Unautorisierter Informationsgewinn	

(Verlust der Vertraulichkeit)	
Unautorisierte Manipulation der Daten (Verlust der Integrität)	
Nachrichten und ihre Inhalte können im nachhinein abgestritten werden (Verlust der Verbindlichkeit)	
Vortäuschen von Client /Datenprovider-Identität (Verlust der Authentizität)	
Unbefugte Funktionsbeeinträchtigung (Verlust der Verfügbarkeit)	

C4. Welche Eigenschaften sollte die Benutzerschnittstelle der Sicherheitsdienste in GBIF-D aufweisen um für Ihre Benutzer akzeptabel zu sein?

☐	Darstellung der Sicherheitsmerkmale von Dokumenten
☐	Konfigurierbarkeit (z.B. An-/Abschalten) von Sicherheitsfunktionen durch Benutzer
☐	Konfigurierbarkeit von Sicherheitsfunktionen ausschließlich durch Administratoren
☐	Möglichkeit zur Einstellung des Sicherheitsgrades (z.B. niedrig, mittel, hoch) durch den Benutzer
Sonstige:	

C5. Welchen Aufwand (Euro oder Personenmonate) verursacht die Einführung / der Betrieb von GBIF-D-Diensten in ihrer Organisation?

z.B.

- Investitionskosten (technische Ausrüstung, Installation, Test, Räume, organisatorische Änderungen (Personal, Verfahren), Abschreibungen usw.)
- Betriebskosten (Mitarbeiter, Kommunikation, Instandhaltung u. Betrieb, Qualitätskontrolle, Personalausbildung usw.)

C6. Welcher Anteil steht hierbei für das IT-Sicherheitsmanagement zur Verfügung? Welcher Anteil könnte zukünftig bereitgestellt werden?

C7. Welche künftigen Entwicklungen / Erweiterungen von GBIF-D würden die Aufwendung zusätzlicher Mittel für Einführung / Betrieb von Sicherheitsdiensten rechtfertigen? z.B. Einführung eines Abrechnungssystems oder Expertenportals

C8. Besteht das Interesse einer technischen Evaluation der Sicherheitsdienste in GBIF-D? z.B. Zertifizierung der sicherheitstechnisch korrekten Funktion durch das Bundesamt für Sicherheit in der Informationstechnik (BSI). Dies könnte aus verwaltungstechnischen oder rechtlichen Gründen (z.B. Signaturgesetz) erforderlich sein. Weitere Informationen unter http://www.bsi.de/literat/faltbl/sichzera.htm

C9. Standards
a) Welche technischen Standards sind für Einführung und Betrieb von GBIF-D-Diensten zu berücksichtigen? z.B. XML, SOAP oder sonstige Übertragungsprotokolle (BioCASE-Protocol, DiGIR, http, ftp usw.)
b) Welche Sicherheitsstandards sind für Einführung und Betrieb von GBIF-D-Diensten zu berücksichtigen? z. B. SSL, PGP, X.509, S/MIME

D. Information

Hilfe Dieser Abschnitt behandelt grundlegende Fragen über die in GBIF-D genutzten Informationen, wie z.B. Datenstrukturen und -flüsse, Kodierungen, (Übertragungs)formate, Sicherung der Qualität der Information im System sowie Identifikation der Ein-/ Ausgabeinformationen.

D1. Welche Arten von Informationen verwenden die von Ihnen angebotenen oder genutzten GBIF-D-Dienste? z.B. : - nach Darstellungsart (Text, Bild, Video usw.) - nach Inhalt (Daten von biologischen Exemplaren/Arten, über Sammlungen, GIS-Daten usw.)

D2. Welchen Schutzbedarf messen Sie diesen Informationen im Einzelnen bei (niedrig, mittel, hoch)?

Information	Schutzbedarf		
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch
	☐ niedrig	☐ mittel	☐ hoch

D3. Datendarstellung (Dateiformate, Kodierungsstandards)

a) **Welche Kodierungsformate werden für die generischen Datentypen benutzt?**
z.B. JPEG, MPEG, XML usw.

b) **Gibt es andere (z.B. biologische) Kodierungsformate für Datenelemente oder sind weitere geplant?**
z.B. ABCD, Darwin Core usw.

D4. In welche Vorgänge sind diese Daten einbezogen (wie und wo werden sie benutzt)?

D5. Datenqualität

Die Erfüllung folgender Sicherheitskriterien trägt zur Verbesserung der Datenqualität bei. Welche dieser Sicherheitskriterien halten Sie innerhalb der GBIF-D-Datenverarbeitungskette (Digitalisieren / Sammeln, Übertragen, Darstellen) für notwendig? Bitte erläutern Sie kurz worin Sie die dadurch erzielte Verbesserung sehen.

☐	Authentizität der Datenquelle bzgl. der Übertragung	
☐	Sicherstellung der Authentizität der Bearbeiter	
☐	Nicht-Manipulierbarkeit der Daten bei der	

	Übertragung	
☐	Nicht-Abstreitbarkeit von Dokumenten	
☐	Vertraulichkeit sensitiver Daten gewährleisten ☐ Komplette Datensätze verschlüsseln ☐ Nur selektive Angaben verschlüsseln (z.B. Lokalitäts-angaben, Namen)	
☐	Zugriff auf Daten nur durch berechtigte Benutzer (Autorisierung).	
☐	Kennzeichnung von Urheberrechten	
Sonstige:		

D6. Welche internen Datenflüsse bestehen zwischen den verschiedenen technischen Komponenten der GBIF-D-Architektur innerhalb Ihrer Institution?

D7. Welche Informationsflüsse von bzw. zu externen Datenprovidern existieren oder sind geplant?

A.2 Ergebnisse

Project	Contacts	Institutions	Responses	% contacts	% institutions
SYNTHESYS	27	19	11	40,74%	57,89%
BioCASE	48	35	17	35,42%	48,57%
GBIF-D	18	16	9	50,00%	56,25%
Total	93	61	37	39,78%	60,66%

Tabelle 1: Statistik über die Rückmeldung auf das 1. Teil der Umfrage

Project	Contacts	Institutions	Responses	% contacts	% institutions
SYNTHESYS	27	19	5	18,52%	26,32%
BioCASE	48	35	10	20,83%	28,57%
GBIF-D	18	16	6	33,33%	37,50%
Total	93	61	21	22,58%	34,43%

Tabelle 2: Statistik über die Rückmeldung auf das 2. Teil der Umfrage

Project	*SYNTHESYS*	*BioCASE*	*GBIF-D*	*Total*
Publication of data	64%	78%	78%	74%
Remote access	36%	61%	56%	53%
Bonding of data sources	9%	44%	44%	34%
Data analysis	45%	22%	22%	29%
Other services	18%	17%	11%	16%
None	9%	0%	0%	3%

Tabelle 3: Angebotene Dienste

Project	*SYNTHESYS*	*BioCASE*	*GBIF-D*	*Total*
Added value by combination of several data sources	73%	89%	89%	84%
Increased name recognition of your data collection	73%	89%	78%	82%
Consistency of the data (expert feedback)	82%	50%	56%	61%
Other goals	0%	28%	33%	21%

Tabelle 4: Vorteile von der Teilnahme am Biodiversitätsnetzwerk

Project	*SYNTHESYS*	*BioCASE*	*GBIF-D*	*Total*
Internal funding	82%	76%	56%	73%
National organisations	45%	76%	67%	65%
International organisations	64%	47%	11%	43%
Commercial organisations	0%	12%	0%	5%
Others	0%	6%	22%	8%

Tabelle 5: Finanzierungsquellen

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Experts	73%	100%	100%	92%
An international public	36%	89%	100%	76%
General public	45%	78%	100%	74%
A national public	18%	89%	100%	71%
Closed user groups	45%	56%	44%	50%
Commercial organisations	18%	39%	33%	32%
Others	0%	11%	11%	8%

Tabelle 6: Publikum für den Dienst

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Yes	91%	83%	67%	82%
Protection against illegal data access	73%	61%	33%	58%
Reliable data sources	45%	61%	11%	45%
Rights management services	36%	67%	11%	45%
Confidentiality of transmitted data	27%	56%	22%	39%
Accounting service	18%	28%	22%	24%
Others	9%	6%	11%	8%

Tabelle 7: Gewünschte Sicherheitsdienste

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Yes	64%	44%	33%	47%

Tabelle 8: Werden sich Sicherheitsdienste positiv auf die Publikationsbereitschaft auswirken

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Data protection and telecommunication acts	36%	50%	11%	37%
Freedom of information	9%	56%	22%	34%
Environment protection laws	9%	50%	22%	32%
Personal and organisational responsibility	18%	33%	0%	21%
unknown	27%	11%	33%	21%
Liability and standardisation	9%	28%	11%	18%
Freedom of privacy	9%	28%	11%	18%
Others	0%	6%	44%	13%
Digital Signature act	0%	6%	0%	3%

Tabelle 9: Rechtliche Restriktionen

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Yes	64%	72%	67%	68%
Internal department	45%	50%	11%	39%
Government, regional or local institutions	45%	33%	22%	34%
Business companies	18%	6%	11%	11%
Others	0%	0%	22%	5%
Certificiation authorities	9%	0%	0%	3%

Tabelle 10: Interne und ausgelagerte schon existierende Sicherheitsdienste

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Firewall	45%	50%	56%	50%
Email service, antivirus, antispam	9%	17%	22%	16%
Others	9%	11%	22%	13%
Digital Signature	18%	6%	0%	8%
Access Control	0%	0%	11%	3%

Tabelle 11: Angewendete Sicherheitsmaßnahmen

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Scientific	**100%**	**100%**	**70%**	**88%**
Institutes	100%	100%	70%	88%
Universities	100%	100%	70%	88%
Biologists	100%	90%	70%	84%
Collectors	80%	90%	70%	80%
Others	*20%*	*10%*	*20%*	*16%*
Companies	*20%*	*0%*	*0%*	*4%*
Consultancies	*20%*	*0%*	*0%*	*4%*
Mapper	*0%*	*0%*	*10%*	*4%*
Education, Schools	*0%*	*20%*	*0%*	*8%*
Government	*0%*	*10%*	*0%*	*4%*
Amateurs	*0%*	*20%*	*0%*	*0%*
Technological	**100%**	**100%**	**40%**	**76%**
Natural scientist	100%	90%	50%	76%
Researcher	100%	90%	50%	76%
Environment protection organisations	80%	90%	40%	68%
Others	*20%*	*10%*	*0%*	*8%*
Agriculture	*20%*	*0%*	*0%*	*4%*
Pharmacy	*20%*	*0%*	*0%*	*4%*
Land-use planners	*20%*	*0%*	*0%*	*4%*
Legal bodies	*0%*	*10%*	*0%*	*4%*
Organisational	**100%**	**100%**	**50%**	**80%**
Direction of museums or institutes	80%	100%	50%	76%
Public offices	80%	70%	50%	64%
Others	*0%*	*10%*	*0%*	*4%*
Regionale Kartierungsstellen	*0%*	*0%*	*10%*	*4%*
NGO	*0%*	*10%*	*0%*	*4%*
Economical	**40%**	**40%**	**30%**	**36%**
Enterprises	40%	40%	30%	36%
Others	*0%*	*0%*	*0%*	*0%*
Other interests	**20%**	**0%**	**0%**	**4%**
Pest organisms	**20%**	0%	0%	4%
Invasive species	20%	0%	0%	4%

Tabelle 12: Vertretene Interessenarten

Project	*SYNTHESYS*	*BioCASE*	*GBIF-D*	*Total*
Confidentiality	**100%**	**80%**	**70%**	**80%**
dispensable	20%	20%	30%	24%
generally important	20%	20%	20%	20%
only for specific data	*80%*	*50%*	*30%*	*48%*
endangered/protected specimen/gems	*20%*	*20%*	*10%*	*16%*
Location of protected plants/Species	*0%*	*10%*	*10%*	*8%*
for consulting firms	*20%*	*0%*	*0%*	*4%*
Integrity	**100%**	**80%**	**70%**	**80%**
dispensable	0%	10%	10%	8%
generally important	100%	50%	50%	60%
only for specific data	*0%*	*20%*	*10%*	*12%*
only by online editing	*0%*	*0%*	*10%*	*4%*
Location of protected plants/Species	*0%*	*10%*	*0%*	*4%*
Reliability/Undeniability	**100%**	**70%**	**70%**	**76%**
dispensable	40%	10%	10%	16%
generally important	60%	60%	40%	52%
only for specific data	*0%*	*10%*	*20%*	*12%*
Annotationen	*0%*	*0%*	*10%*	*4%*
Location of protected plants/Species	*0%*	*10%*	*0%*	4%
Authenticity	**100%**	**70%**	**70%**	**76%**
dispensable	0%	0%	10%	4%
generally important	80%	60%	30%	52%
only for specific data	*20%*	*10%*	*30%*	*20%*
Location of protected plants/Species	*0%*	*10%*	*0%*	*4%*
Availability	**100%**	**50%**	**70%**	**68%**
dispensable	20%	0%	0%	4%
generally important	40%	40%	60%	48%
only for specific data	*40%*	*20%*	*10%*	*20%*
Location of protected plants/Species	*0%*	*10%*	*0%*	*4%*

Tabelle 13: Relevanz der IT-Sicherheisziele für das Biodiversitätsnetzwerk

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Unauthorised data modification	**100%**	**60%**	**20%**	**52%**
Loss of reliability	40%	30%	20%	28%
Additional time and staff effort	0%	0%	10%	4%
Service become unusable for sensitive data	40%	20%	0%	16%
Corrupted Data	0%	10%	0%	4%
Unauthorised gain of information	**60%**	**60%**	**20%**	**44%**
Loss of ownership	20%	30%	0%	16%
Service become unusable for sensitive data	20%	10%	10%	12%
Disclosure of location about protected/endangered specimen	20%	0%	0%	4%
Minor consequences	0%	20%	0%	8%
Loss of scientific reputation	0%	0%	10%	4%
Deniability of messages and contents	**100%**	**30%**	**20%**	**40%**
Loss of rights to act legally	20%	0%	0%	4%
Loosing the trust of the data sources	0%	10%	20%	12%
Additional time and staff effort	0%	0%	10%	4%
Loss of Data quality	20%	0%	0%	4%
Minor consequences	20%	0%	0%	4%
Service become unusable for sensitive data	20%	10%	0%	8%
Pretention of a Client/Provider identity	**80%**	**20%**	**10%**	**28%**
Loss of ownership	20%	10%	0%	8%
Loss of reliability	20%	10%	10%	12%
Additional time and staff effort	0%	0%	10%	4%
Service become unusable for sensitive data	40%	10%	10%	16%
Unauthorised service restriction	**60%**	**20%**	**20%**	**28%**
Reduced service level	20%	0%	0%	4%
Dimishing the provider image	0%	10%	10%	8%
Additional effort for making the service available again	20%	0%	10%	8%
Minor consequences	20%	0%	0%	4%
Service become unusable for sensitive data	0%	10%	0%	4%

Tabelle 14: Konsequenzen beim Verlust der IT-Sicherheitgrundwerte

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Configuration (e.g. on/off) of security features by administrators	100%	80%	50%	72%
Presentation of the security attributes of documents	80%	60%	50%	60%
Possibility to select a security level by the user	0%	30%	30%	24%
Configuration (e.g. on/off) of security features by the user	0%	20%	10%	12%
Others	0%	0%	0%	0%

Tabelle 15: Gewünschte Visualisierung / Kofigurierbarkeit von Sicherheitsmerkmalen

Project	SYNTHESYS	BioCASE	GBIF-D	Total
None	0%	10%	0%	4%
5% - 10%	0%	30%	0%	12%
< 20 %	0%	10%	0%	4%
30%	20%	0%	0%	4%

Tabelle 16: Anteil der Mittel für die Entwicklung von Sicherheitsdiensten

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Expert portal	60%	20%	0%	20%
Copyrights	40%	10%	0%	12%
Accounting system	20%	10%	0%	8%
Online data editing	0%	0%	10%	4%
Quality control of exact observation data	0%	0%	10%	4%
GBIF Integration	0%	10%	0%	4%
Funds applying	0%	10%	0%	4%

Tabelle 17: Weitere Entwicklungen / Einsätze, die ein sicheres System fordern

Project	SYNTHESYS	BioCASE	GBIF-D	Total
Interested	40%	40%	40%	40%
No	20%	40%	0%	20%
Eventually	0%	10%	20%	12%
Unknown	20%	0%	0%	4%

Tabelle 18: Interesse an einer Evaluation der Sicherheitsdienste

Project	SYNTHESYS	BioCASE	GBIF-D	Total
BioCASE-Protocol	40%	50%	60%	52%
XML	20%	40%	40%	36%
http	40%	40%	20%	32%
DiGIR	20%	30%	20%	24%
SOAP	0%	30%	10%	16%
https	20%	10%	0%	8%
ABCD-Schema	20%	0%	10%	8%
Databases (MS Access)	0%	20%	0%	8%
SQL	20%	0%	0%	4%
ftp	0%	10%	0%	4%
Recorder 2002	0%	10%	0%	4%

Tabelle 19: Eingesetzte technische Standards

Project	SYNTHESYS	BioCASE	GBIF-D	Total
SSL	40%	20%	10%	20%
Unknown	20%	20%	0%	12%
S/Mime	20%	10%	0%	8%
Any	20%	0%	0%	4%
PGP	0%	10%	0%	4%

Tabelle 20: Eingesetzte Sicherheitsstandards

Project	*SYNTHESYS*	*BioCASE*	*GBIF-D*	*Total*
Authenticity of data provider during transmission	**60%**	**90%**	**40%**	**64%**
Ownership, Attribution	40%	10%	0%	12%
Integrity of data during transmission	**80%**	**70%**	**40%**	**60%**
Quality Control	40%	20%	0%	16%
Only authorised data access	**80%**	**70%**	**30%**	**56%**
Only for experts or sensitive data	40%	0%	0%	8%
Usage within terms and conditions of use	20%	0%	0%	4%
Securing of the authenticity of the originator	**60%**	**60%**	**40%**	**52%**
Ownership	20%	10%	0%	8%
Warranty of the confidentiality of sensitive data	**60%**	**70%**	**30%**	**52%**
Access only for approved recipients	20%	0%	0%	4%
Complete encryption	***40%***	***10%***	***10%***	***16%***
Partial encryption	***20%***	***50%***	***30%***	***36%***
Marking of copyrights	**60%**	**50%**	**50%**	**52%**
Visible attribution	20%	0%	0%	4%
Undeniability of documents	**40%**	**20%**	**30%**	**28%**
spurious data prevention	20%	0%	0%	4%

Tabelle 21: Wichtigkeit der IT-Sicherheitgrundanforderungen für die Datenverarbeitung