

Public Worldwide Distributed Plant Seed Genebank

Peter Fargaš

<https://informatik-handwerk.de>

0160 / 93 42 37 52

Eingereicht bei [DLR Projektträger](#)

Bundesanzeiger: BAnz AT 01.08.2016 B6

Dokument Versionen & Errata

2016 Oktober 27

- o offizielle "Citizen Science" Projektskizze
- o pdf erhältlich auf Anfrage

2016 Oktober 28

- o Korrigierte Formatierung
- o Segmentierung der Kapitel "Idee und Ziele" zur Erhöhung der Lesbarkeit
- o [pdf download](#)

[Link zur autoritative Quelle](#)

≡ Inhalt

[Dokument Versionen & Errata](#)

1. [Selbstdarstellung](#)
 2. [Kurzzusammenfassung](#)
 3. [Ausführliche Vorhabenbeschreibung](#)
 - 3.1 [Idee und Ziele](#)
 - 3.2 [Akteure und Zielgruppen](#)
 - 3.3 [Forschungsdesign, Zeitplanung](#)
 - 3.4 [Daten: Umgang und Qualität](#)
 - 3.5 [Kommunikationsmaßnahmen](#)
 - 3.6 [Evaluation](#)
 4. [Eigeninteresse/Eigenanteil](#)
 5. [Nachhaltigkeit, Übertragbarkeit](#)
 6. [Budgetschätzung](#)
- [Quellenangabe](#)

Selbstdarstellung

Ich habe theoretische Informatik studiert und bin als Entwickler über 15 Jahre professionell in der Industrie beschäftigt. Ich habe mich auf rigorose Prototypisation (algebraisch abgeschlossene Räume und deren Unifikationen), Sammlung und Auswertung von Datenreihen (Energieindustrie) sowie Web (Design, User Interfaces) spezialisiert. In meiner Freizeit betreibe ich Forschungen zur Benutzeroberflächen-Entwicklung und zur Systemtheorie und Systemdynamik.

Ich verfüge über relevante Kontakte zur Universität Leipzig, zur Comenius Universität Bratislava (Slowakei) sowie dem Sublab-Hackerspace in Leipzig und anderen unabhängigen Machern.

Zur Zeit bin ich selbständig.

Ziele des Vorhabens und Kurzzusammenfassung des Vorhabens

Erstellung einer weltweiten, verteilten Genbank für Pflanzensamen, Initiierung einer öffentlichen Bewegung zur Sicherung der Biodiversität.

Idee und Ziele: Darstellung der wissenschaftlichen Fragestellung und des bürgerwissenschaftlichen/Citizen Science Ansatzes

Vorhandene Landschaft

Genbanken, genomische Bibliotheken, DNA-Bibliotheken sind Institutionen die sich ex situ der Erhaltung von Biodiversität und der Bereitstellung des dafür notwendigen Materials widmen. Die grundlegende Idee wurde in einfacher Form bereits seit der frühen Landwirtschaft angewendet. In letzten Jahrzehnten kamen neue Gründe und Herausforderungen dazu.

Der Zustandswandel unser Habitat und die Ursachen dafür sind allgemein bekannt^[1]. Je mehr die Industrie an Gewicht gewinnt, desto mehr gilt es auch die Konsequenzen in Betracht zu ziehen. Zu den mittlerweile berechenbaren Langzeit-Konsequenzen von klassischen Industrien kommen die rein theoretisch geschätzten Konsequenzen von High-Tech-Industrien (d.h. experimentellen Industrien). Ein Beispiel für eine Technologie, deren versteckte Gefahren einfach zu verstehen sind - die Sterilisation-Strategie, "Technology Protection System" (TPS)^[2]. Konkret bei diesem Beispiel wird auch inkonsistente Argumentation deutlich.

Sowohl die Industrie als auch wissenschaftliche Institutionen beschränken sich angesichts der überwältigenden Aufgabe größtenteils auf Kulturpflanzen. Eine stark komprimierte Liste von Genbanken, die meiner Meinung nach aber zumindest die Hierarchie der Genbanken offenbart, gibt es bei Wikipedia^[3]. Die Situation in Deutschland wird u.a. von der "Genbank Bayern Arche" und dem Bundesinformationssystem für Genetische Ressourcen dargestellt^{[4],[5]}.

System Vorschlag

Die technologische Ausrüstung, zuhause eine kleine Genbank zu administrieren, besitzt heutzutage innerhalb "des westlichen" Lebensstandards jede und jeder. Leute wünschen sich essenziell zu sein, Sinn in ihr Leben zu bringen, an wichtigen Aufgaben zu partizipieren, "hands-on"-Erfahrungen und Fehler machen zu dürfen, Einsichten in hochwertige Verfahren und Kenntnisse für sich zu gewinnen, großartigen Hobbys und Freizeitbeschäftigungen nachzugehen, Teil von Kommune und Bewegung zu sein - das wäre ein Beiprodukt und gleichzeitig Argument für die Durchführbarkeit des Vorhabens.

Eine redundante, verteilte Resource Description Framework-Datenbank (RFD), die sich in die vorhandene Landschaft von vernetzten Informationen^[4] nahtlos einfügt, versehen mit einem Web-Interface, das der Verwaltung und dem sozialem Vernetzen dient, ist der einzige fehlende Teil, um ein solches System in die Existenz zu rufen.

Seitens der Nutzer handelt es sich um ein Sammelhobby - etwas wie Philatelie, Numismatik, das Einfangen von Pokèmons oder die Beschäftigung mit einem Stickeralbum, allerdings ist hier ein tieferer Sinn present. Das Hobby benötigt keine finanzielle Mittel, die obere Grenze vom Wert der eigenen Sammlung ist einfach zu sprengen und der Einsatz von Gamification und Geocaching-ähnlichen Aktivitäten eröffnet neue Dimensionen - beim Spaßfaktor und sozialer Vernetzung, aber auch bei der Wissensvermittlung. Nutzergerecht zu sein, ist hier ebenso wie bei den anderen Variationen möglich: von Kinderalben zum Durchblättern mit Vergrößerungsglas, zu speziellen Verpackungen von einzelnen Akzessionen und einem Passivmonitoring der mittels Radio-frequency-identification-Technologie (RFID) ordentliche Qualitätssicherung erzielt. Das meiste aber machen bereits einfachen Methodiken aus: wie Lichtdichteverpackung, Lagerung im Tiefkühlfach oder Beilage von Trockenmitteln.

Tauschen untereinander und die Bereitstellung für wissenschaftliche Institutionen sind Grundbestandteile des Hobbys und zusätzliche Motivationsfaktoren. Die vorher erwähnte Gamification und das Geocaching sind nicht nur als Spaßfaktoren, sie dienen auch dazu, die Vervollständigung der Genbank und die Balance ihres Inhalts zu sichern. Eine weitere wichtige Eigenschaft solcher verteilten Systeme ist es, dass sie sich problemlos Mehrfachbelegungen leisten können und somit eine der robustesten Systeme sind, die überhaupt entwickelt werden. Wie schon zum Ausdruck kam, sind die Anforderungen von verteilten Systemen an deren Teile extrem niedrig, und deren Heterogenität ist sogar vom Vorteil. Die Anwendung von statistischen Verfahren, die zur Qualitätsestimation von Akzessionen und Korrektheit der Daten dienen, geben dem allem eine feste Grundlage.

Als Daten- und Sozial-Netzwerk ist das System eine spezialisierte Datenbank mit dedizierten REST- und öffentlichen Web-Zugängen. Massendaten-Anfragen und Filter-/Suchfunktionalität werden für wissenschaftliche Institutionen freigegeben. Für die breite Öffentlichkeit steht eine gängige (dennoch spezialisierte) Benutzeroberfläche zur Verfügung. Von den notwendigen und weniger klassischen Elementen würde ich gerne Visualisierung, die Darstellung von Geodaten, die Wissenshierarchie der Nutzer und das gegenseitige Bemühen um Korrektheit der Daten hervorheben (vgl. Wiki- Prinzip o.ä.). Das ganze Geschehen kann auch als einmaliges Sozialexperiment gesehen werden und ist daher selbst gut geeignet für eine solche Auswertung. Anonymisierte Nutzer-Daten sollten hierfür bereitgestellt werden.

Akteure und Zielgruppen: Wer ist am Forschungsprozess beteiligt?

- breite Öffentlichkeit, Pflanzenzüchter, Kleingärtner
- Gärtner, Förster und Waldverwaltung, Naturschutzgebietsmanagement
- Schulen und Erziehungsinstitute
- Wissenschaftliche Institutionen, Bibliotheken und Datensammlungscentren
- Samenhändler
- Gene-Fonds und Institutionen für die Absicherung der menschlichen Existenz

Arbeits- und Zeitplanung: Darstellung des Forschungsdesigns, der Methoden und des Ressourcenbedarfs

Forschungsdesign/Methodik:

Der System-Entwurf erfüllt folgende Voraussetzungen:

- einfache und direkte Motivation der Nutzer (vertrauter Ansatz, Gamification)
- Robustheit (verteilt System)
- Skalierbarkeit, unbegrenzte Kapazität (verteilt System)
- niedriger Wartungsaufwand (kollektive Aufgabe)
- Systematik der Arbeit mit Daten (heterogener Ansatz, statistische Verfahren)
- Balancierung von Resultaten (Lückenerkennung, Gamification als Call-for-action)

Eine genauere Beschreibung findet sich im Abschnitt [Idee und Ziele - System Vorschlag](#) (Systemdesign) und im Abschnitt [Angaben zum geplanten Umgang mit Daten - Qualitätssicherung der Daten](#) (Qualitätsmanagement).

Technologien

- Datenspeicherung: SQL, tripplestore
- Datenformate: XML/JSON, RDF, Dublin core - in Abhängigkeit von Partnern
- Protokolle: RESTful, GUNet
- Architektur und Server Seite: Cloud oder dezentralisiert
- Client-Seite und Benutzeroberfläche: java, Browser
- Visualisierung: d3
- Datenauswertung: statistische Verfahren, Ranking

Controlling des Arbeitsprozesses:

- A. Vorbereitung auf vorhersehbare Ereignisse: Feststellungen des zukünftigen Personal-, Material-, Wissens- und anderen Bedarfes
- B. Resilienz gegen unvorhersehbare Ereignisse: Backup und andere Strategien
- C. Vorbeugung von vorhersehbaren Fehlentscheidungen: Revision von Entscheidungen, Plausibilitätsprüfungen
- D. Transparenz und Beihilfe ähnlicher Projekte: Erfahrungsberichte und Bereitstellung von Resultaten der einzelnen Schritte, Dokumentation
- E. Qualitätsmanagement von (Teil-)Resultaten
- F. Koordination und Absprache mit Partnern, Zeit-Management
- G. Revision und Optimierung von Arbeitsabläufen

Grober Arbeitsplan:

Phase 1 (0%-20%) :

1. Öffnung der Kommunikationskanäle
 1. Peer-Institutionen, Datenzentren
 2. Feststellung der Interessen auf institutionellen Ebenen
2. Planungsverfeinerung und Konzeption
3. Budget-Gliederung und bürokratische Strukturierung
4. Technologisch-infrastrukturelle Vorbereitungen
5. Erste Blind-Skizzen
 1. Interne Daten-Repräsentation und deren Funktionalität
 2. System-Design
 3. Nutzungs-Szenarien
 4. Web- und Mobile-App-Interfaces

Phase 2 (15%-30%):

1. Workshops und Verhandlungen mit Partnern
 1. Feststellen von deren Personal- und Budget-Kapazitäten
 2. Feststellen von deren Nutzungs-Szenarien und Anforderungen
2. Analyse von bestehenden Datensammlungen mit Fokus auf gute technologische und administrative Anbindung
 1. Feststellung der technologischen Möglichkeiten und Grenzen
3. Entwicklung der Strategien zur Abschätzung von öffentlichen Bedürfnissen
4. Erstes öffentliches Erscheinungsbild
5. Investorensuche und Design von anderen Möglichkeiten, um den dauerhaften Betrieb zu gewährleisten (selbsttragende technologische Infrastruktur z.B.)

Phase 3 (30%-35%):

1. Technologische Entscheidungen
2. Entscheidungen zu externen Prozess- und Arbeitsabläufen
3. Verhandlungen zu Daten-Schreibzugang, externe Qualitätssicherung und Absorption von Daten
4. Durchführung der Strategien zur Abschätzung des öffentlichen Bedürfnisses
5. Openourcing/Closed Source-Festsetzung, Lizenzentscheidungen
6. Skalierbarkeitsszenarien

Phase 4 (35%-50%):

1. Runder Tisch und offizielle Begutachtung in Präsenz allen Beteiligten
2. Verfestigung der allgemeinen Roadmap und Zielsetzung, von internen Meilensteinen und Terminzusagen nach außen

Phase 5 (45%-65%):

1. Aufsetzen der technologischer Infrastruktur
2. Entwicklung und Begutachtung von Daten-Anbindungen
3. Entwicklung von hausinterner Datensatzspeicherung und der Interfaces
4. Prototyp von Massendaten-Zugängen
5. Prototyp und Case-Studies von User Interfaces
6. Entwurf der statistischen Verfahren

Phase 6 (55%-75%):

1. Außenerscheinungsbild
 1. Strategien zur Einbeziehung der Öffentlichkeit
 2. Corporate Identity
 3. Planung einer Werbekampagne
 4. PR Strategien
2. Implementation von öffentlichem Web- und App-Interface
3. Continuous Feedback-Sammlung und Einarbeitung

Phase 7 (70%-90%):

1. Datenbefüllung
2. Belastungsproben
3. Privates und semi-öffentliches Testing
4. Fehlerbehebung, Tuning
5. Wiki, Foren und andere Methodiken für Kontakt mit und Feedback der Öffentlichkeit

Phase 8 (90%-100%):

1. öffentlicher Start
2. Notfall-Einsätze
3. Stabilisierung der Lage

Post-Phase (100+):

1. Dauerhafte Feedback-Auswertung und Einarbeitung
2. Erhaltung und Wartung der Infrastruktur

Angaben zum geplanten Umgang mit Daten: Erhebung von Daten, Datenqualität, Datensicherung und Speicherung von Daten, Zugänglichkeit von Daten, Zitation von Daten, etc.

Arten von Daten

- persönliche Daten und mit Nutzer-Identität zusammenhängende Daten
 - Persönliche Daten, Nutzer-Daten
 - Kommunikation und Wissensaustausch
 - Buchführung von Aktivität (Log)
- auf Akzessionen bezogene Daten und dem Zweck dienende Daten
 - Klassifikation und Quantifikation von Material
 - Qualitätsschätzungen
 - Fotografien, Geodaten
 - Daten zur Einbettung von Material in weiteren Kontext

Datenquellen

- externe Datenzentren
 - Einbringung eigener Grundsätze, die es zu befolgen gilt
- von Nutzern bereitgestellte Daten
 - dreifach getrennt in öffentliche, on-request und systeminterne Daten
 - nach gängigen Datenschutzrichtlinien behandelt

Wege zur Gewährleistung von Qualität

- Einzelne Datensätze
 - Selbstkontrolle: Bewertung und Belohnung von Vollständigkeit
 - auf sozialer Basis: Soziales Netzwerk mit gegenseitige Bewertung
 - auf Qualifikation basierend: Wissenshierarchie der Nutzer
 - Datenqualität: gegenseitige Kontrolle und Korrektur
 - Materialqualität: Bewertung von erhaltenen Samenbestellungen
 - Prozessqualität: Auskunft zur Verpackung und anderen "Hardware"-Konfigurationen, z.B. RFID-Passiv-Monitoring(Temperatur, Feuchtigkeit), Trockenmittelbeilage, Vakuumierung, Kühlung, Licht.
- Datensätze im Kontext
 - statistische Verfahren zur Abschätzung von Qualität der gelagerten Akzessionen
 - statistische Verfahren zur Abschätzung der Datenaktualität
 - statistische Verfahren zur Abschätzung der Verlässlichkeit der Nutzer
 - Priorisierung der Pflanzenarten in Abhängigkeit zur geographischen Lokalität, bereits vorhandenem Samenbestand und dessen geschätzte Verfügbarkeit und Qualität

Auswertung von Daten

- Statistische Verfahren zur Abschätzungen der Qualität und tatsächlichen Verfügbarkeit
- Visualisierungen, basierend auf d3.js Framework
- wird durch externe Institute durchgeführt

Zugänglichkeit

- (semi-)öffentliche Schnittstelle, welche die Daten im gängigen RDF-Format präsentiert, wird Institutionen und dem inneren Kreis der Interessenten zur Verfügung gestellt
- frei zugängliches Web-Interface nach Registrierung
- öffentliche Visualisierung von Daten und regelmäßige Berichte
- anonymisierte Nutzerstatistiken auf Anfrage

Kommunikationsmaßnahmen: Wie soll der Forschungsprozess öffentlich begleitet werden und wie werden die Ergebnisse vorgestellt?

Speziell am Anfang sollte eine Werbekampagne die Entstehung popularisieren und Interesse daran wecken, hauptsächlich durch zielgruppenbezogene Informationskanäle (Kleingartenverwaltungen; relevante Vereine; Information solcher Eltern, die besondere Vorstellungen von den Freizeitaktivitäten ihrer Kinder haben könnten; thematisch-komplementäre Foren und Internetkommunen etc.) aber auch Pressenachrichten bei Zeitschriften wie GEO sind gut vorstellbar.

Ergebnisse, im Sinne von regelmäßigen Berichten - denn es gibt kein Zeitpunkt, an den man sagen könnte, der Prozess sei abgeschlossen - werden mit d3.js Framework visualisiert. Daraus folgend können wissenschaftliche Berichte erstellt werden. Der direkte Zugang zu Daten wird auch gewährleistet.

Bestellungen von Samensamples durch einzelne Partizipanten zu ermöglichen, ist eine der Kernaufgaben des Systems. Diese Samen können zum direktem Einsatz kommen wie auch digitalisiert werden, sei es als high-resolution Aufnahmen, Dissektion oder als DNA-Sequenzierung. Besondere Befunde von Nutzern würden in übergeordnete Genbanken übernommen.

Evaluation: Benennung von internen Evaluationskriterien und -indikatoren für die erfolgreiche Umsetzung des Projekts

- Der Ausgangspunkt für eine Evaluation ist die Größe der partizipierenden Öffentlichkeit, das Wachstum an trivialen Datensätzen und Samenmaterial wie auch die basale soziale Aktivität. Dies könnte als passives Ausprobieren/Rauschen der Präsenz im System gesehen werden.
- Aktive Teilnahme wäre das Eingehen auf Gamification und die Beherrschung von üblicherweise selten gelagerten Samen, die Vervollständigung der Daten und Befunde usw.
- Ein wichtiger Indikator ist die Steigerung der Qualität der Lagerung.
- Herauskrystallisieren von Community-Leaders, aktive Teilnahme von Experten und gegenseitige Hilfe bei der Datenerfassung und die Richtigkeit von Daten sind ebenso einfach erfassende Merkmale.
- Die komplette und dichte Deckung von allen bewachsenen Gebieten ist das ultimative Ziel.

Darstellung des Eigeninteresses/Eigenanteils

Mein Großvater, Vít Bojnanský, war Biologe und aktiv an der agrikulturellen Entwicklung der Slowakei im letzten Jahrhundert beteiligt. Unsere Familie ist in Besitz von ca. 7000 Akzessionen aus den gesamten Karpaten, von denen ca. 4500 ihren Platz in der Publikation "ATLAS OF SEEDS AND FRUITS OF CENTRAL AND EAST-EUROPEAN FLORA"^[6] fanden. Wir stehen aktuell in Verhandlung, diese Daten für eine öffentliche Nutzung freigeben und somit selbst die ersten großen Datensätze liefern zu können. Zum Vergleich: laut <http://www.genbank-bayern-arche.de/> wäre die Genbank die mein Großvater erstellte, die drittgrößte von Deutschland.

Konzeptuell ist die Verbindung von beliebten Freizeitbeschäftigungen mit einer essenziellen Aufgabe eine großartige Idee und einzigartig. Die Kosten sind amortisiert auf Null und der Ansatz präsentiert eine neue Dimension im Bereich von Freizeitbeschäftigungen. Diese Einzigartigkeit kombiniert mit der Handfestigkeit des Augenblickes des Starts wird bestimmt auch in soziologischen Kreisen Interesse wecken. Außerdem bin ich überzeugt davon, dass die Kernaufgabe des Projektes nur als kollektives Werk zu bewältigen ist.

Ich habe natürlich auch professionelles Interesse an der Leitung eines solchen Projekts.

Nachhaltigkeit, Übertragbarkeit

Der dauerhafte Durchfluss von Samen durch die Nutzer erstellt einen Puffer, aus denen sich Nutzer und Institutionen jederzeit bedienen können. Wertvolle Akzessionen können übernommen werden, um ihnen einen höheren Sicherheitsstandard zu geben.

Die damit zusammenhängenden Daten werden intern in gängigen Datenaustauschformaten gespeichert und sind somit bereit zur Weiterverarbeitung. Massendatenzugänge, spezialisierte Suchverfahren und Datenquellen-Korrelations-Werte werden bereitgestellt.

Eine breite Öffentlichkeit bekommt Wissen und Anregungen zu Verläufen von qualitativ hochwertigen Prozessen und zu Qualitätsstandards.

Budgetschätzung

Das Projekt wurde auf eine Laufzeit von drei Jahren hin konzipiert, eine vereinfachte Variante kann in zwei Jahren fertiggestellt werden.

Personalkosten		
1 Person Vollzeit		90,000 €/Jahr
1 Person Halbezeit Aushilfe (geschätzt, nach Bedarf)		18,000 €/Jahr
Informationen		
Literatur und andere Quellen		500 €/Jahr
Schulungen		1,500 €/Jahr
Extern durchgeführte professionelle Aufgaben		
Verwaltung von Finanzen, doppelter Buchführung, etc.		4,000 €/Jahr
Technologische Infrastruktur		3,000 €/Jahr
Werbung und PR		2,500 € einmalig
Infrastruktur		
Bürobedarf, Telekommunikationskosten		1,000 €/Jahr
Arbeits- und Prozess-Infrastruktur		2,000 €/Jahr
Reisekosten		1,500 €/Jahr
Material		
Grafiken & Flyer, Visitenkarten, Druck		2,500 € einmalig
Laufende Kosten		n/a

Projektkosten	121,500 €/Jahr + 5,000 € einmalig	
50% Anteilfinanzierung	60.750 €/Jahr + 2,500 € einmalig	
=====		
2-Jähriges Szenario	ca. 120,000 €	"De-minimis"-Beihilfen
3-Jähriges Szenario	ca. 185,000 €	"De-minimis"-Beihilfen

Quellenangabe

1. http://www.genbank-bayern-arche.de/about_genbanken.html#situation
2. http://www.gmo-compass.org/eng/safety/environmental_safety/173.environmental_safety_stopping_spread_foreign_genes.html
3. https://en.wikipedia.org/wiki/Seed_bank#Facilities
4. <http://www.big-flora.de/partner-und-big-kandidaten/>
5. http://www.genbank-bayern-arche.de/about_genbanken.html#home
6. <https://link.springer.com/book/10.1007/978-1-4020-5362-7>