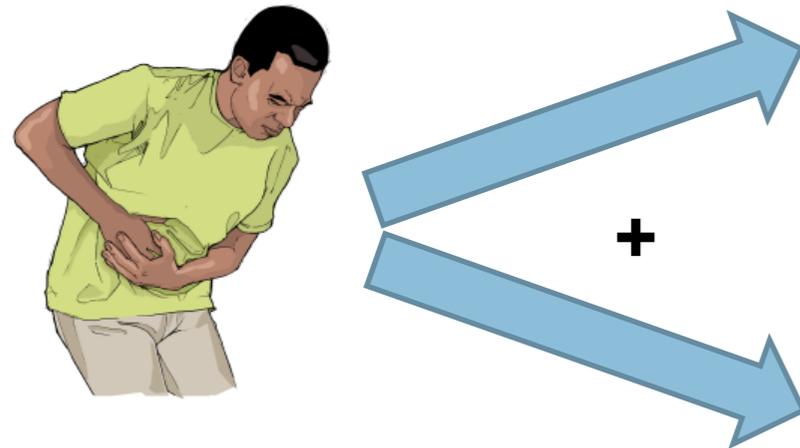
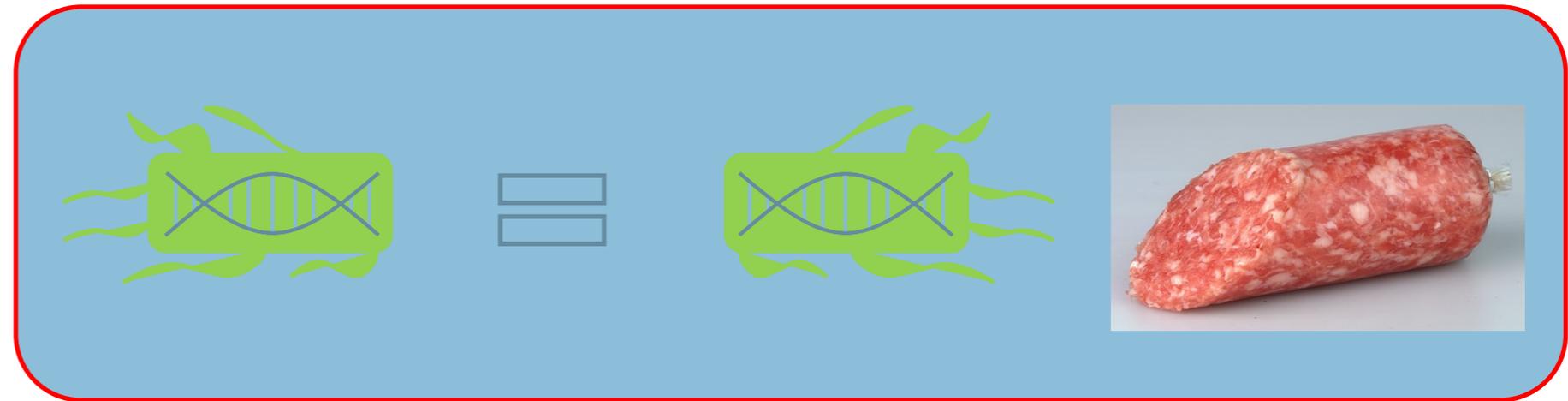


Whole Genome Sequencing Datenmanagementkonzepte: Nationale und Europäische Entwicklungen

Carlus Deneke, BfR

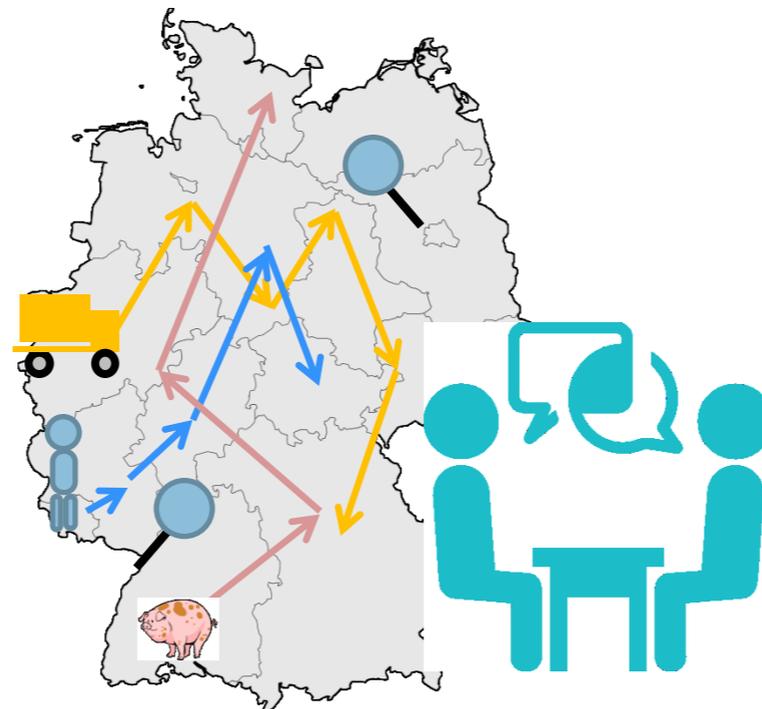
Untersuchung von lebensmittelbedingten Krankheitsausbrüchen

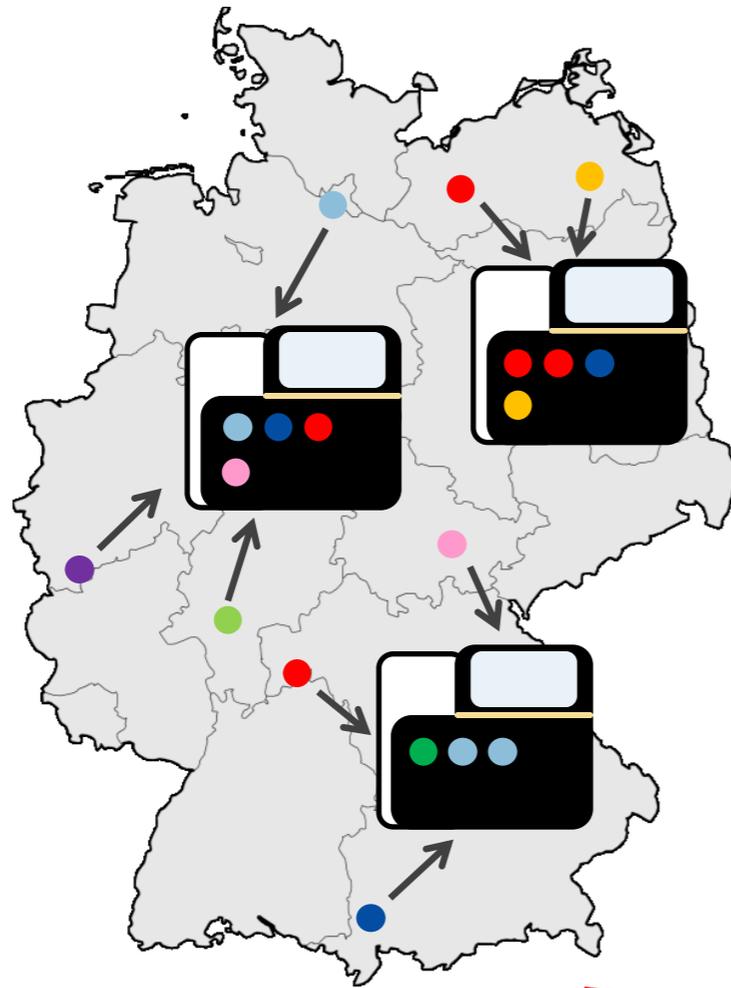
Labordiagnostik: Typisierung des Ausbruchsstamms



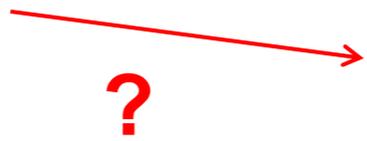
Epidemiologie und Rückverfolgung

- Befragungen
- Warenketten
- Inspektionen
- ...

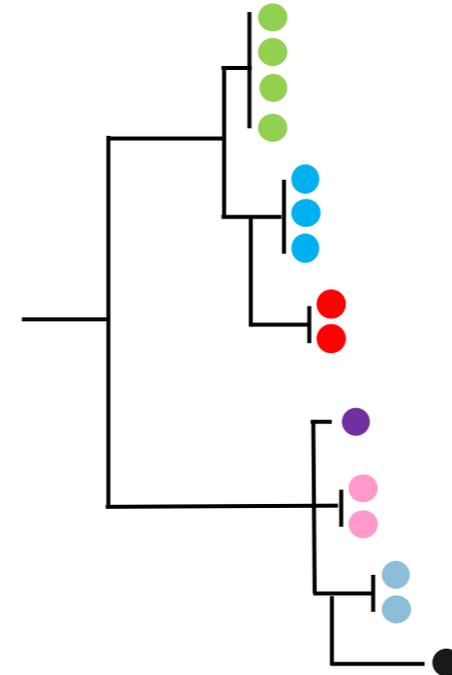




Gesamtgenomsequenzierung der Isolate in unterschiedlichen Einrichtungen



Fragestellung:
Wie können wir mit verschiedenen generierten Sequenzdaten eine gemeinsame Analyse durchführen?



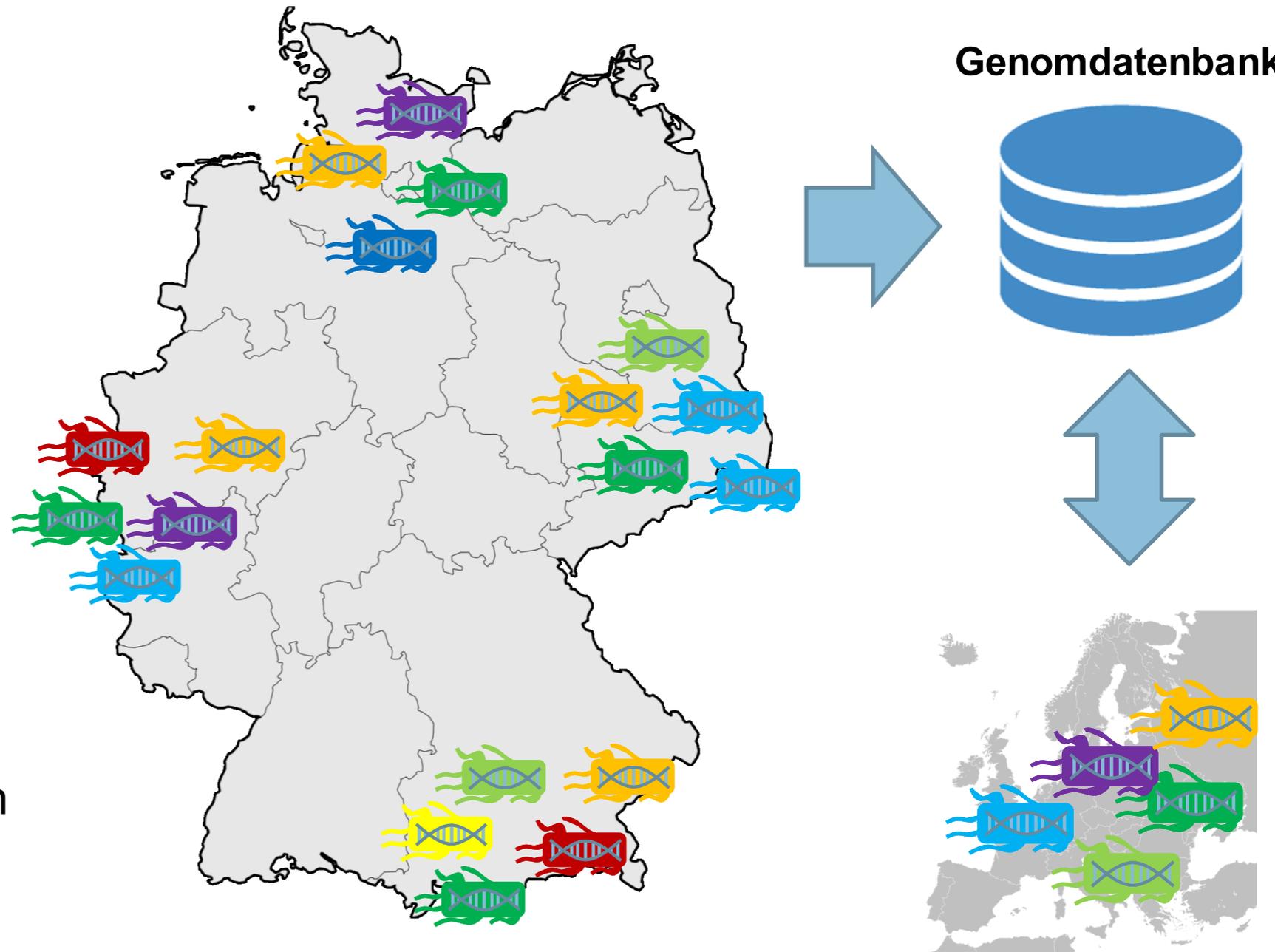
Status-quo der WGS-gestützten Ausbruchsbekämpfung

- Keine gemeinsame und sektorübergreifende Datenbank der Typisierungsdaten in Deutschland (und Europa)
 - Anlassbezogener Austausch von Genomdaten, z.B. bei größeren Krankheitsausbruchsgeschehen
- Keine Harmonisierung der bioinformatischen Analysemethoden zur Typisierung
 - Z.B. Nutzung von verschiedenen kommerziellen und open-source Software Programmen
 - Bewertung Aufgrund von Gen-zu-Gen (cgMLST) oder SNP Untersuchungen
 - Lokale Analysen auf nicht-vollständigen Datenmengen

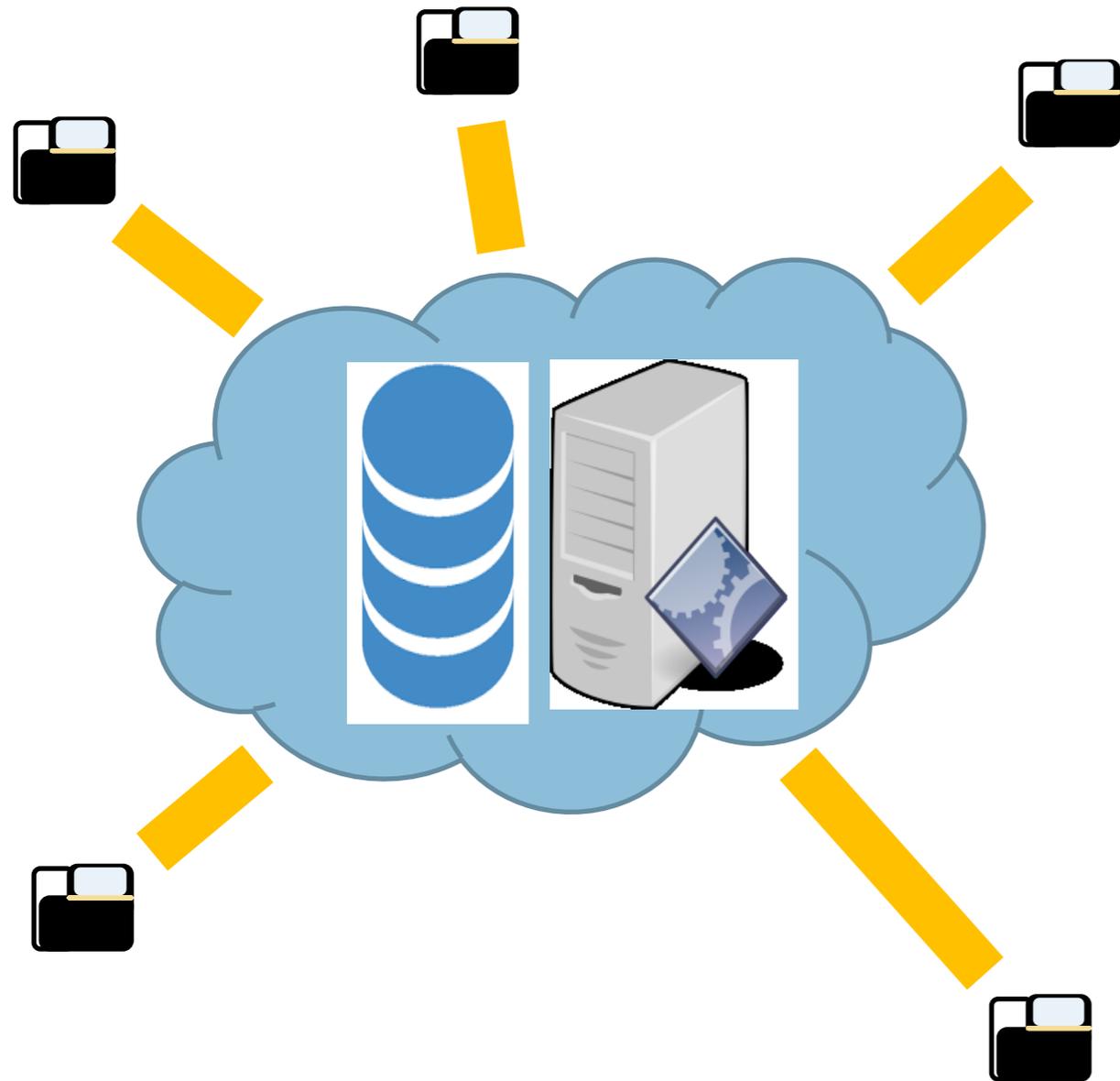
-> „Anlassbezogener Genomvergleich“

Anforderungen an prospektive Genomanalysen

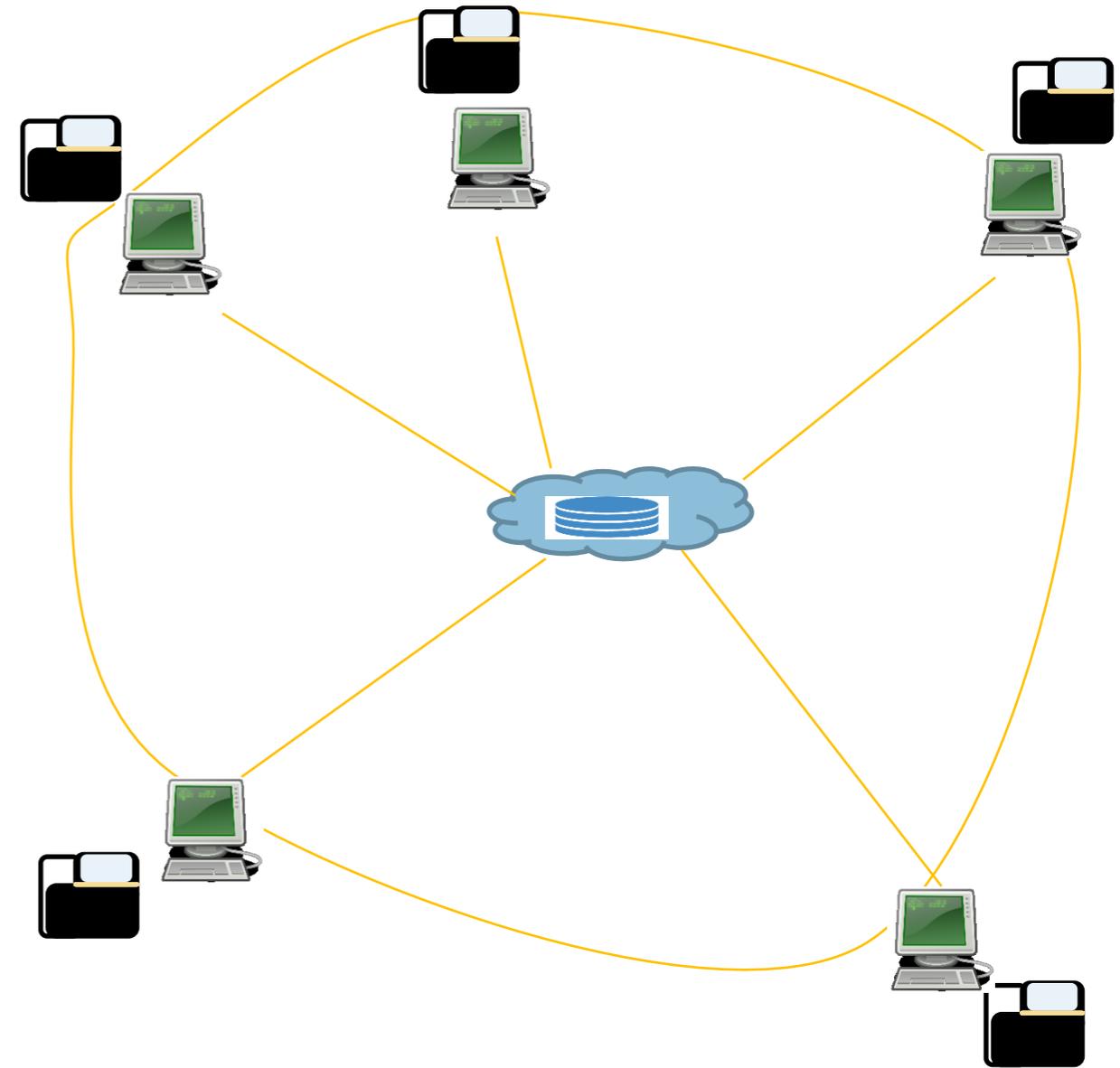
- Einrichtung einer zentralen Datenbank:
 - Spezialisiert auf WGS Daten
 - Großer Speicher
 - Verfügbarkeit
 - Datenauthentizität und –integrität
 - Datensicherheit
- Standardisierung der Analysemethode
 - Automatisierte Clusteridentifizierung und Informationsweitergabe
- Abgleich mit europäischen / internationalen Ausbruchsgeschehnissen



Lösungsansätze



Austausch der Rohdaten und zentrale Analyse

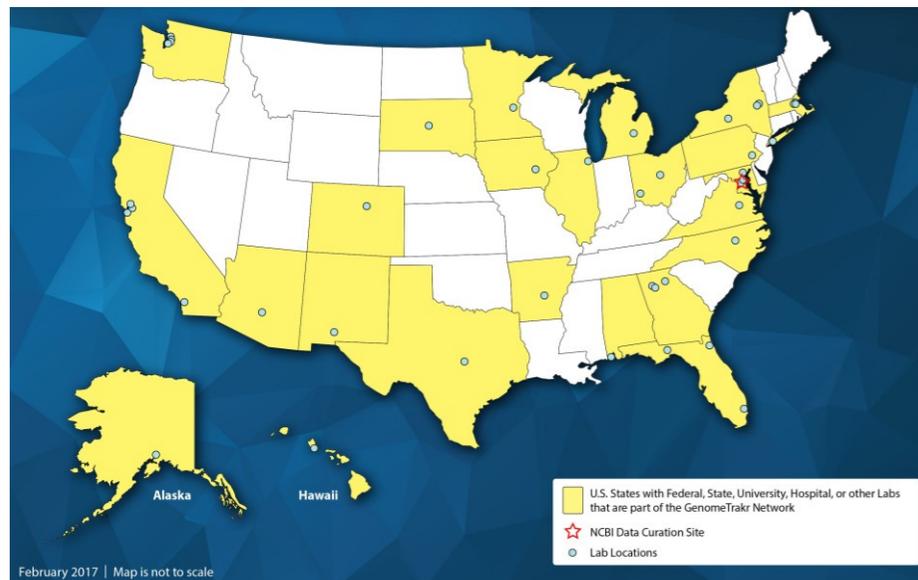


Dezentrale Analyse und zentrale Ergebnisdatenbank

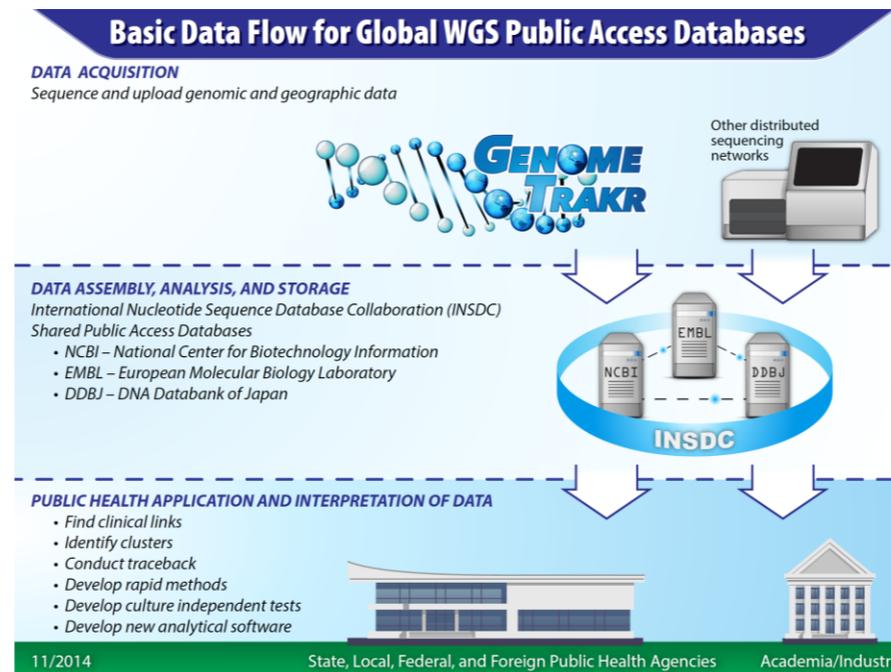
GenomeTrakr Programm in den USA

GenomeTrakr Programm in den USA

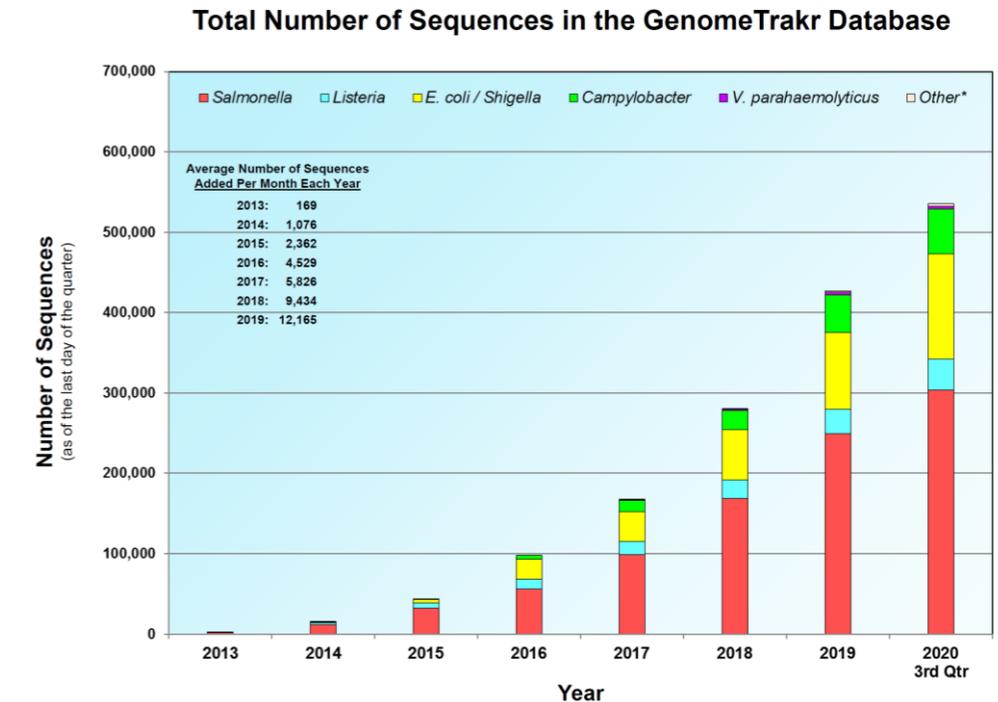
- Dezentrale Sequenzierung über USA verteilt (CDC, FDA, Labore der Bundesstaaten)
- Zeitnahe und **zentrale** Datensammlung der Rohdaten in der öffentlichen Datenbank der NCBI (*OPEN DATA* Ansatz)
- Bioinformatische Analyse lokal durch CDC, FDA UND **zentral** (und öffentlich) durch NCBI
- Fortlaufende, gemeinsame Bewertung von Ausbrüchen durch CDC und FDA
- Viele International Partner: z.B. Public Health England
- <https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-network>



<https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-network>



<https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-fast-facts>



<https://www.fda.gov/food/whole-genome-sequencing-wgs-program/genometrakr-fast-facts>

GenomeTrakr Programm in den USA

- NCBI Portal zur bioinformatischen Analyse
<https://www.ncbi.nlm.nih.gov/pathogens/>
- ~40,000 *Listeria monocytogenes* Sequenzen verfügbar und analysiert
- ~3000 SNP Cluster
- Analyseergebnisse (Molekulare Typisierung) stehen der ganzen Welt zur Verfügung

Implementation of Nationwide Real-time Whole-genome Sequencing to Enhance Listeriosis Outbreak Detection and Investigation FREE

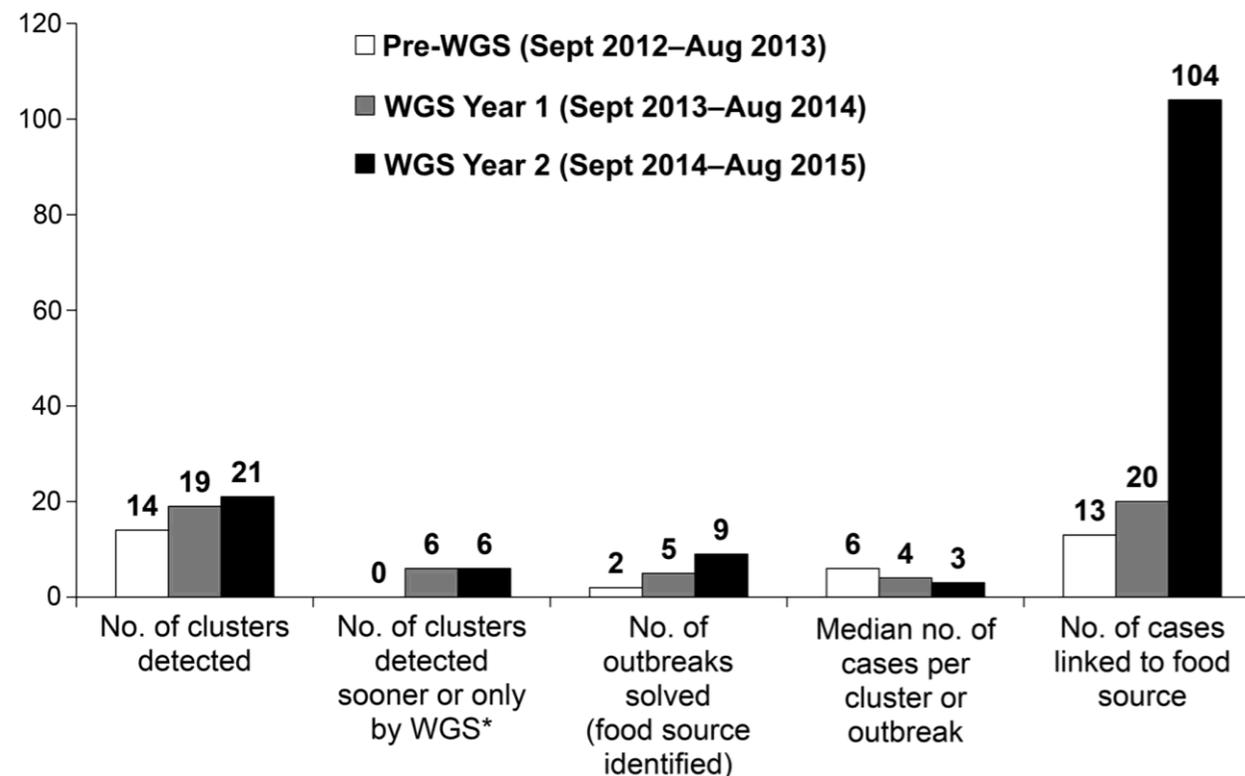
Brendan R. Jackson ✉, Cheryl Tarr, Errol Strain, Kelly A. Jackson, Amanda Conrad, Heather Carleton, Lee S. Katz, Steven Stroika, L. Hannah Gould, Rajal K. Mody ... [Show more](#)

Clinical Infectious Diseases, Volume 63, Issue 3, 1 August 2016, Pages 380–386,

<https://doi.org/10.1093/cid/ciw242>

Published: 18 April 2016 [Article history](#) ▼

Clin Infect Dis, Volume 63, Issue 3, 1 August 2016, Pages 380–386, <https://doi.org/10.1093/cid/ciw242>



FDA/CDC Real Time *Listeria* Project

Further integrating the marriage of clinical, food, and environmental pathogen analysis, FDA has partnered with CDC in an effort to **sequence every clinical, food, and environmental isolate** of *Listeria monocytogenes* collected in the U.S. CDC is primarily sequencing clinical samples while FDA and the state laboratories are sequencing food and environmental samples. The genomic sequences and corresponding collection information for the samples are **publicly available** via the NCBI website.

AFFL-Projektgruppe „Lebensmittelbedingte Krankheitsausbrüche, analytischen Möglichkeiten und Folgen“

AFFL-Projektgruppe „Lebensmittelbedingte Krankheitsausbrüche, analytischen Möglichkeiten und Folgen“

- Aktiv seit Oktober 2019, Vorsitz Bayern
- **Auftrag:** Erarbeitung von **Vorschlägen zu Koordinierungsaufgaben** im Falle länderübergreifender Krankheitsausbrüche, analytische Möglichkeiten, **Zuständigkeiten von Bund und Länder**
- In Diskussion:
 - WGS-Datenmanagementkonzept vorgestellt durch das BfR
 - Koordinierung zwischen Bund-Länder Behörden (einschließlich des RKI) bei Ausbruchsklustern
- Erste Ergebnisse nicht vor Sommer 2021 zu erwarten

Lösungsvorschlag zum WGS Datenmanagement

Genomdaten



Öffentlich

Kontinuierliche, zeitnahe Veröffentlichung von Genomsequenzierungsdaten zentral auf ENA/NCBI Datenbank durch Labore der Länder und des Bundes

- Einschließlich minimaler Basisinformation: Staat, Isolationsjahr, Quelle (Lebensmittel/Tier/Umwelt)
- Standardisierte Auswertung verfügbar („Pathogen Detection Pipeline“)

Metadaten



Nicht-öffentlich

Übermittlung von datenschutzrelevanten **Metadaten** zum Isolat in eine einzurichtende **nicht-öffentliche Datenbank**

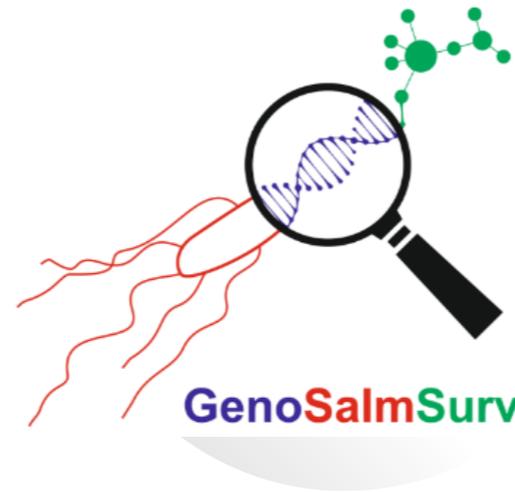
- Zu übermittelnde Metadaten: Übermittler, Original-Isolatnummer, Probennahmedatum, Matrix, Bundesland, Erregervorbefund, Ansprechpartner
- Informationen zum Betrieb verbleiben bei den Länderbehörden

Vorteile der vorgeschlagenen Datenbanklösung

- Prospektive und automatisierte Sequenzanalyse führt zu schnelleren Erkennung von überregionalen Ausbrüchen
- Veröffentlichung/Speicherung der Sequenzdaten kostenneutral möglich, keine Entwicklungsarbeit von speziellen Sequenzdatenbanken notwendig
- Standardisierte Analysemethode zur Ermittlung der Isolateigenschaften (Genomähnlichkeit und Antibiotikaresistenzprofile) bereits verfügbar
- Alle Labore haben Zugang zu Genomdaten und können bei Bedarf eigene Untersuchungen durchführen
- Der Ansatz entspricht der europaweiten open-data Policy
- Lösung entspricht weltweiten Bestrebungen (GMI Initiative) zur Identifikation internationaler Cluster
- Öffentliche WGS Daten erlauben den Abgleich von Eigenkontrollen der Unternehmen

GenoSalmSurv Projekt

gefördert durch:  Bundesministerium
für Gesundheit



→ Aufbau einer integrierten genombasierten Surveillance für Salmonellen

Zusammenführung der genombasierten Typisierungsdaten mit epidemiologischen Daten, sowie Daten der Lebensmittel- und Veterinärbehörden

GenoSalmSurv: Ablauf

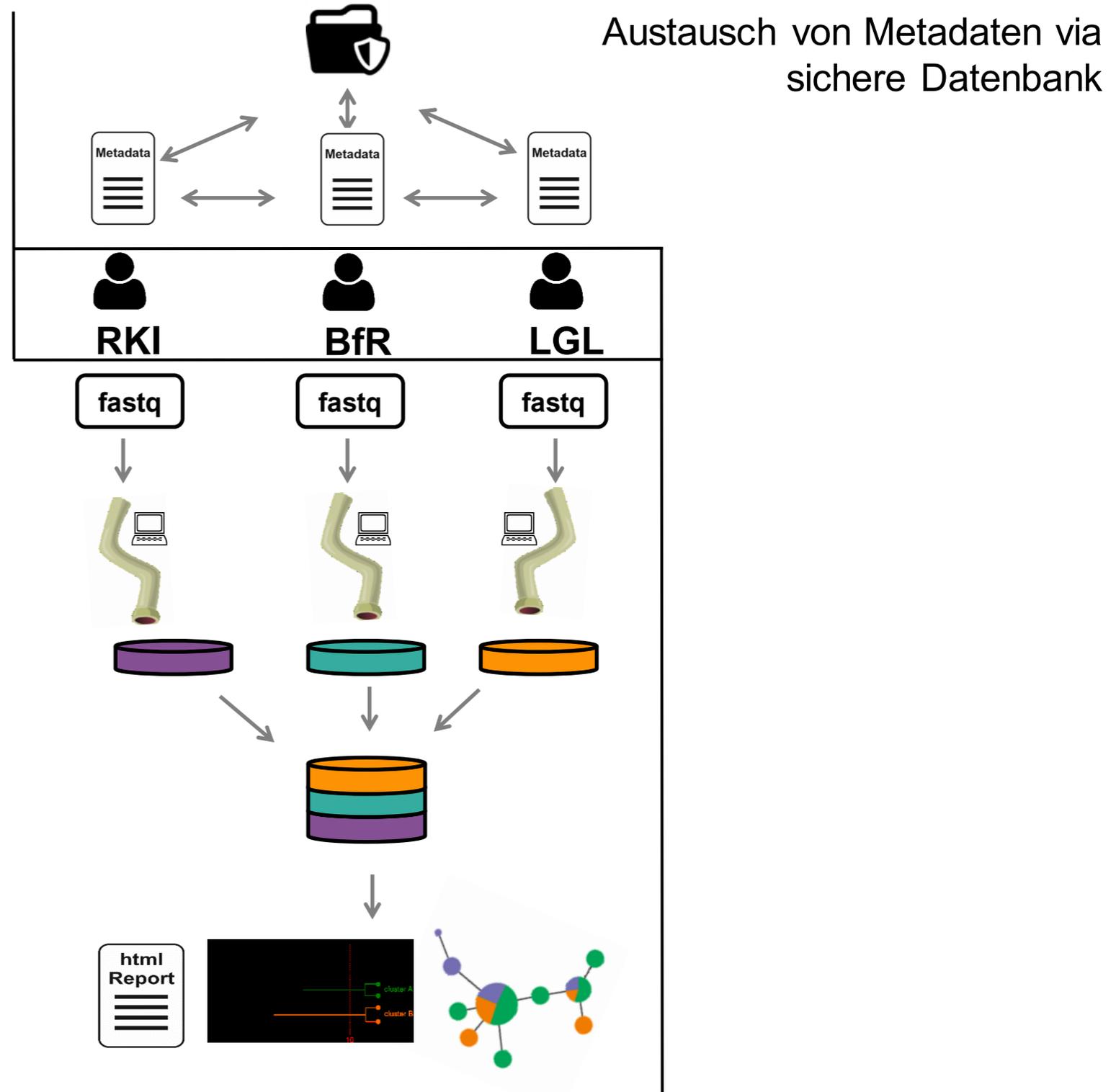
Generierung von Sequenzdaten

Lokale bioinformatische Analyse (Assemblierung & cgMLST Analyse)

Zusammenführung der Ergebnisdaten (Allelprofile)

Generierung von Cluster-Reports

- + Dezentrale Analyse
- + Austausch der Ergebnisdaten



GenoSalmSurv: Realtime-Erprobung

- Sequenzierung und bioinformatische Auswertung von allen einkommenden Isolaten
- Serovare: Enteritidis, Typhimurium, Infantis, Choleraesuis
- Validierung von Pipelines und Clusterdetektion
- Erprobung des Datenaustauschs und der Kommunikation
- Gemeinsame Clusterbewertung
- Ablauf in zwei Phasen zu je 6 Monaten:

1. Phase

Mai – Oktober 2020

RKI: ~500 Isolate

BfR: ~500 Isolate

LGL: ~100 Isolate

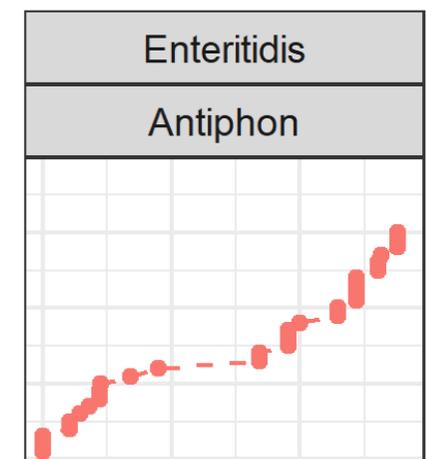
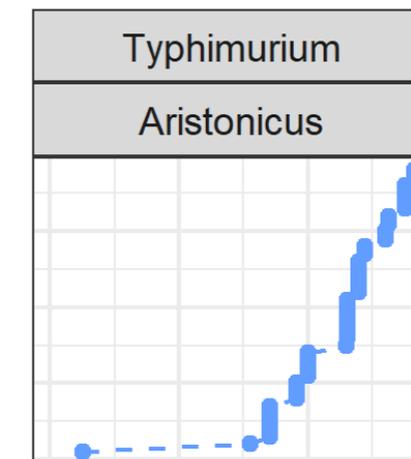
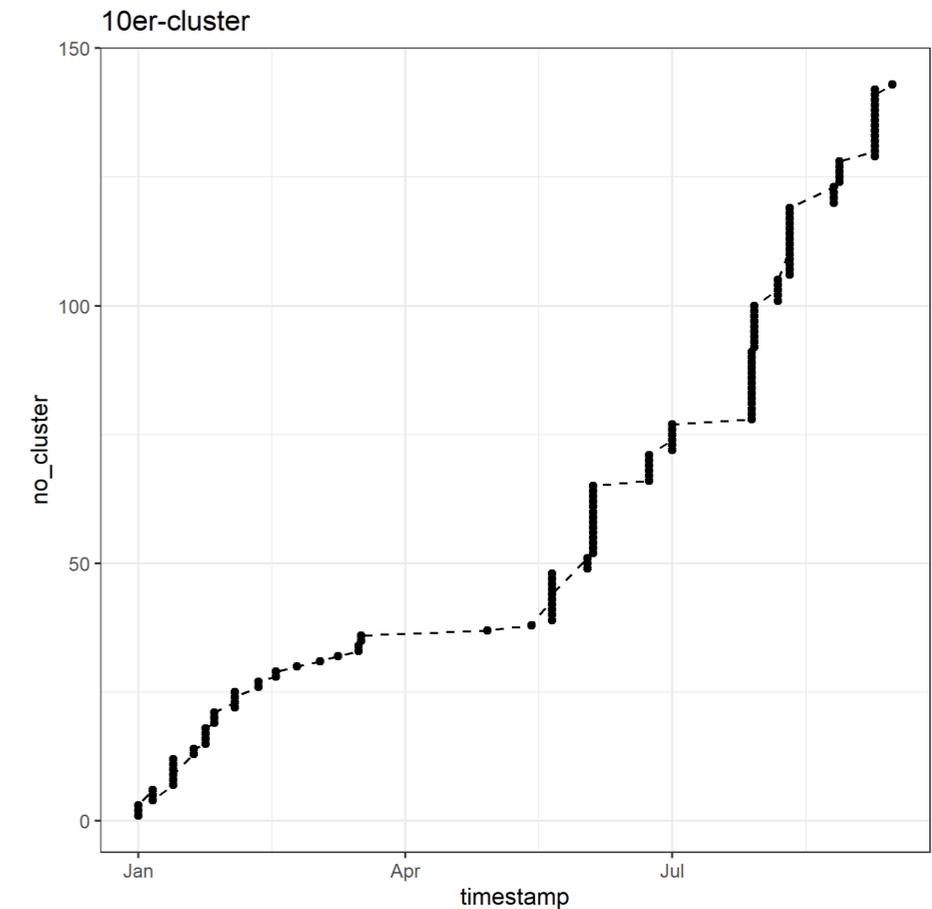
2. Phase

geplant Februar – Juli 2021

RKI: ~500 Isolate

BfR: ~500 Isolate

LGL: ~100 Isolate



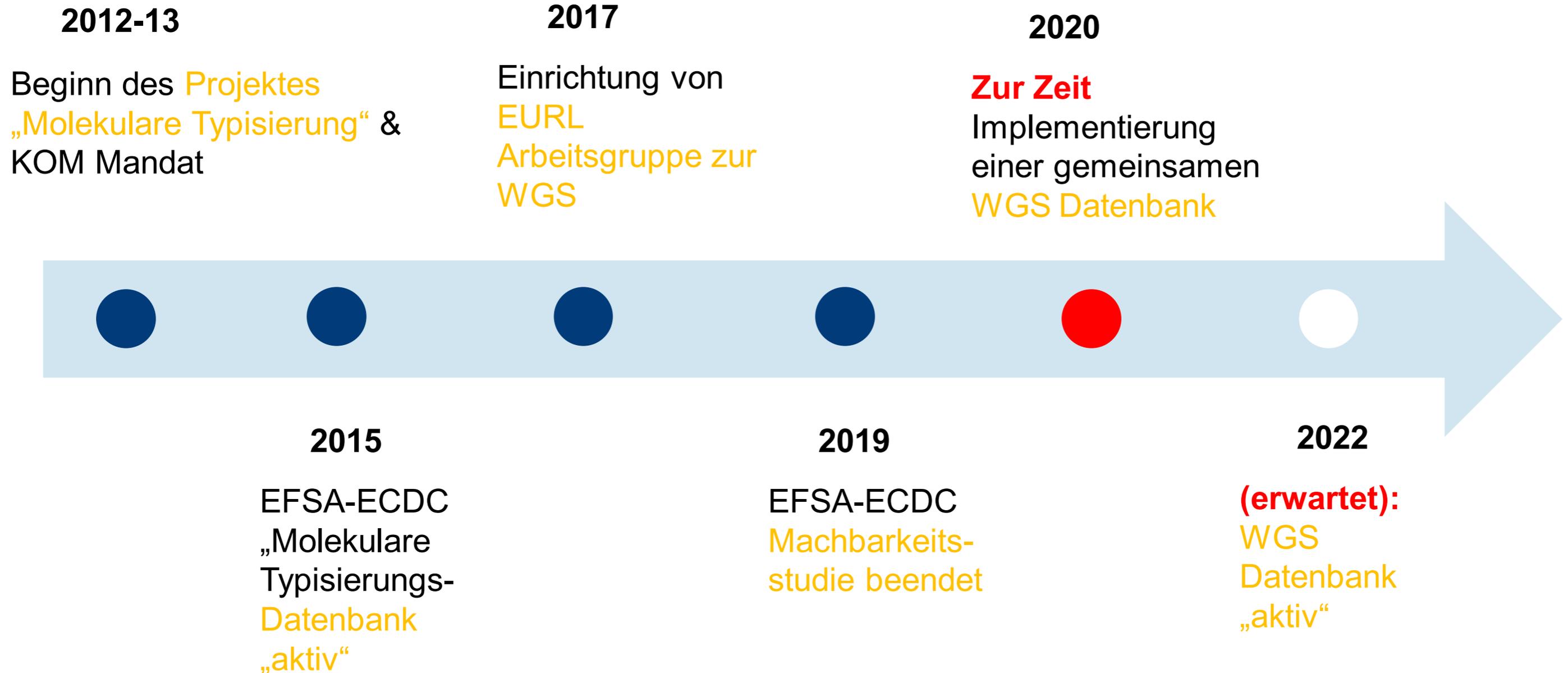
GenoSalmSurv: Zusammenfassung

- Lokale Analysen funktionieren
- Ergebnisaustausch funktioniert zeitnah nach Sequenzierung
- Clustererkennung erfolgt ebenfalls zeitnah
- Viele neue Cluster werden erkannt
- Bewertung von Clustern, Priorisierung der Rückverfolgung herausfordernd

- Methode auch schon erfolgreich für Listeria getestet

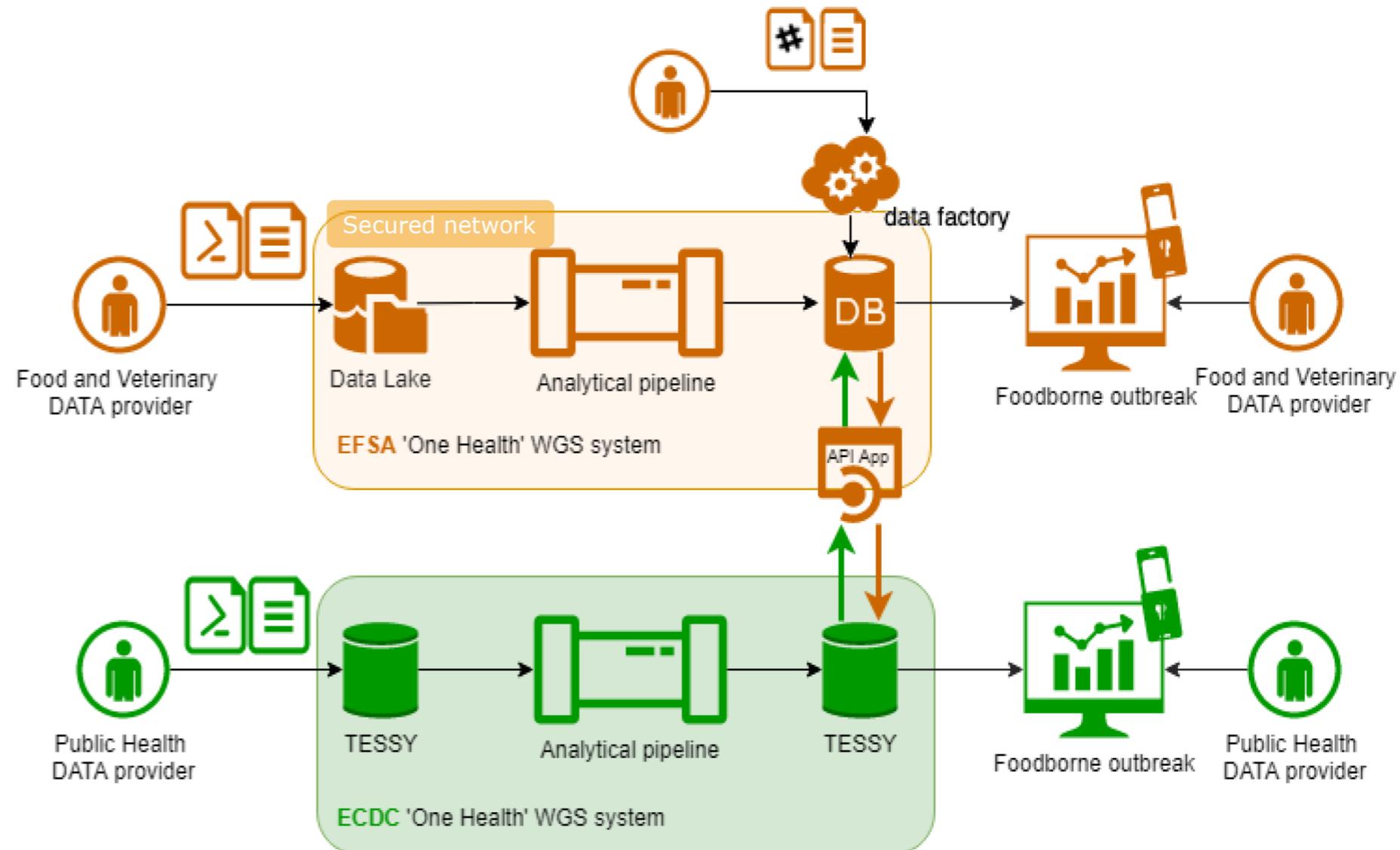
Europäische Aktivitäten zur Errichtung von molekularen Datenbanken (EFSA-ECDC)

Europäische Aktivitäten zur Errichtung von molekularen Datenbanken (EFSA-ECDC)



Vorgesehene Datenbank-Lösung der EFSA-ECDC

- Open-source Lösung basierend auf **cgMLST**
- Dezentralisierte Ergebniserzeugung, Übertragung der Ergebnisse in zentrale Datenbank
- Rohdaten können beim Besitzer bleiben
- Einbindung von ENA/SRA Daten
- Harmonisierung und Austausch der Ergebnisdaten mit ECDC



Quelle: A. Rizzi. EFSA State of play on NGS and future perspectives in the food safety area. „Science meets Policy“ conference 25. September 2020. ISS, Italien

Jede Menge Herausforderungen.....

- Potential von WGS unbestritten für die Überwachung und Monitoring von Gesundheitsgefahren und in epidemiologischen Untersuchungen von Erregern, aber....

Viele Herausforderungen kommen damit auf uns zu

Organisatorisch

- Gesetzgebung
- Regulation
- Infrastruktur
- Finanzierung
- Ökonomie
- Priorisierung

Kulturell

- Sektor-übergreifend
- politisch
- sprachlich
- Hemmnisse zur Veränderung

Technisch

- Datenschutz
- Metadatenformat
- Datenspeicherung
- Datentransfer
- Rechenleistung
- Qualitätsmanagement

Wissenschaftlich

- Sequenzierplattform
- Bioinformat. Tools
- Nachfolgende Analysen
- Training und Kapazität
- Harmonisierung

Quelle: Whole genome sequencing for foodborne disease surveillance. Landscape paper. WHO, 2018

Zusammenfassung

- Verschiedene nationale und internationale Bestrebungen zur Schaffung einer sektor- und länderübergreifenden, gemeinsamen Bewertung der molekularen Typisierung
- **Dezentrale** Lösungsansätze mit Austausch von Ergebnis-Daten basierend auf **open-source** Bioinformatik Software
 - In der Erprobung für Salmonellen im GenoSalmSurv Projekt
 - Geplant für 2022 für die EFSA Datenbank
- **Zentrale** Lösungsansätze mit zentraler, **open-data** WGS Datenbank
 - GenomeTrakr (USA und internationale Partner)
 - AFFL-PG (Bund-Länder Projektgruppe)
- Die Verschiedenen Lösungsansätze sind integrierbar!

Danke für Ihre Aufmerksamkeit

Dr. Carlus Deneke

Bundesinstitut für Risikobewertung

Max-Dohrn-Str. 8-10 • 10589 Berlin

Tel. 030 - 184 12 - 0 • Fax 030 - 184 12 – 99 0 99

bfr@bfr.bund.de • www.bfr.bund.de