



Erste Resultate und Ausblick

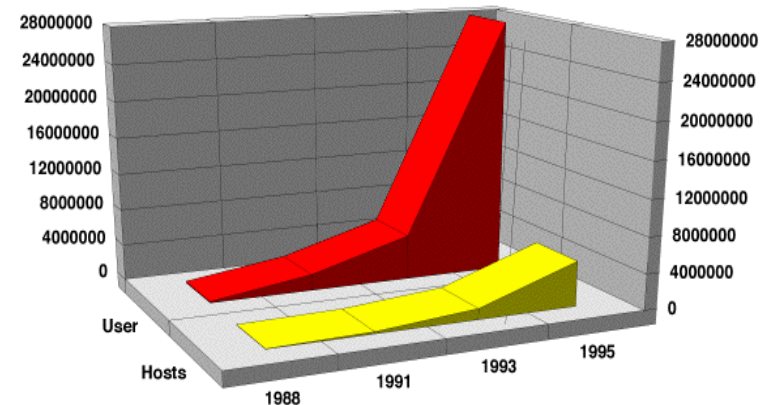
**RDA Deutschland Treffen
November 2014**

**Peter Wittenburg
Max Planck Gesellschaft**

**research data sharing without barriers
rd-alliance.org**

Kurzer Blick zurück

- Internet – Austausch von Botschaften/Daten durch die Vernetzung von Computern
 - viele Senior Wissenschaftler in D sehr zögernd
 - die Wiss/Publications-Maschine funktionierte
 - brauchte 15-20 Jahre, um Fahrt aufzunehmen
 - Durchbruch durch TCP/IP Übernahme und WWW
 - Impakt für Forschung wurde erst später klar
 - Effizienzgewinn & neue Optionen
 - heute sind Internet & WWW nicht wegzudenken
- ~1993 ein weiterer Boost – TCP/IP akzeptiert, Email etc.
- in der Wissenschaft
- WWW/Mosaic



Trends

- sind wir uns einig?
 - Daten-Vielfalt, -Menge und –Komplexität nehmen zu
 - wir sprechen über Methoden/Infrastrukturen in ca. 5 Jahren, wenn manuelles Arbeiten an einzelnen Dateien/Dokumenten immer mehr zu einer Nische wird (Disziplin-Unterschiede)
 - daten-getriebene Forschung (Big Data), um kompetitive Resultate präsentieren zu können
 - Kollaborationen & Datenaustausch über Länder- und Disziplingrenzen hinweg werden zu einer Selbstverständlichkeit

Arbeit mit Daten

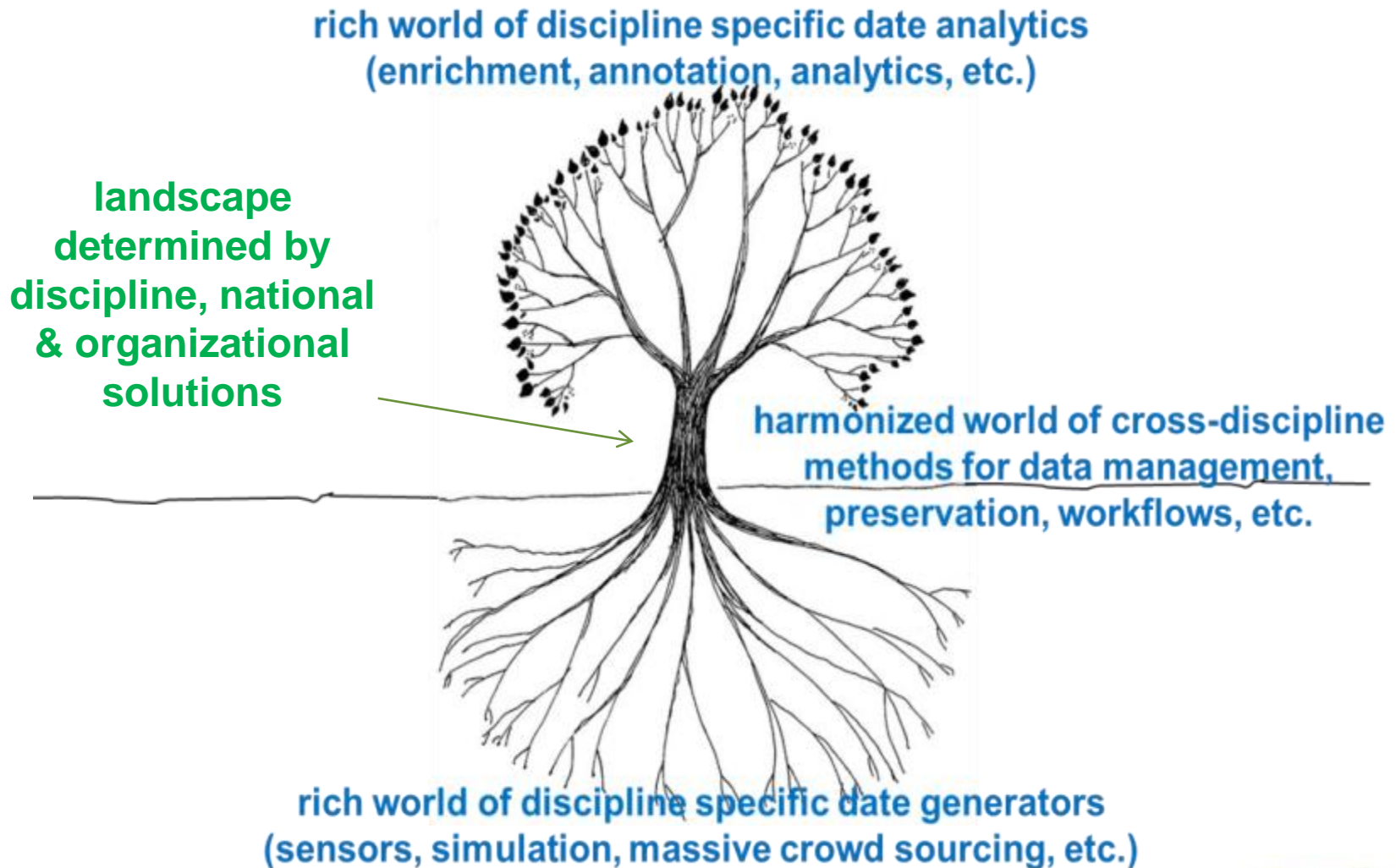
- einige Institute betreiben daten-intensive Forschung - haben jedoch relativ viel Geld und Spezialisten
- kennen diverse Leuchtturm-Projekte mit durchdachten Infrastrukturen
- Informatik hat oftmals (theoretische) Lösungsvorschläge
- aber ...
 - das umfassende **Integrieren** von Daten verschiedener Quellen ist zu aufwendig (EUDAT, DataOne, CLARIN, GEO-China, etc.)
 - **Notlösungen regieren**: Integration auf File oder “Cloud-Objekt” Ebene – dabei Ignorieren aller “logischer” Informationen (Metadaten, PIDs, Rechte, Relationen, etc.), viele ad-hoc Scripts, manuelles Operieren, etc.
 - Konsequenz: eine **ineffiziente** und **nicht-reproduzierbare Wissenschaft** – gegen alle Grundsätze

Umfrage Resultate

- ca. 120 Interviews mit
 - verschiedene Senior Wissenschaftler einig:
 - wir müssen Datenorganisation und Prozeduren ändern – können so nicht weitermachen
 - dies ist momentan riskant und es fehlen die Daten-Experten zur Umsetzung
 - man vermisst klare Orientierungshilfen
 - Workshop: <https://europe.rd-alliance.org/Content/Events>
 - Survey: <https://europe.rd-alliance.org/Content/NewsRoom.aspx?id=221&Page=1&Caption=0!3>

gegen weiter Legacy-Daten in großem Stil, da Richtlinien für eine gute Datenorganisation und Software fehlen

DM ist Disziplin-Unabhängig



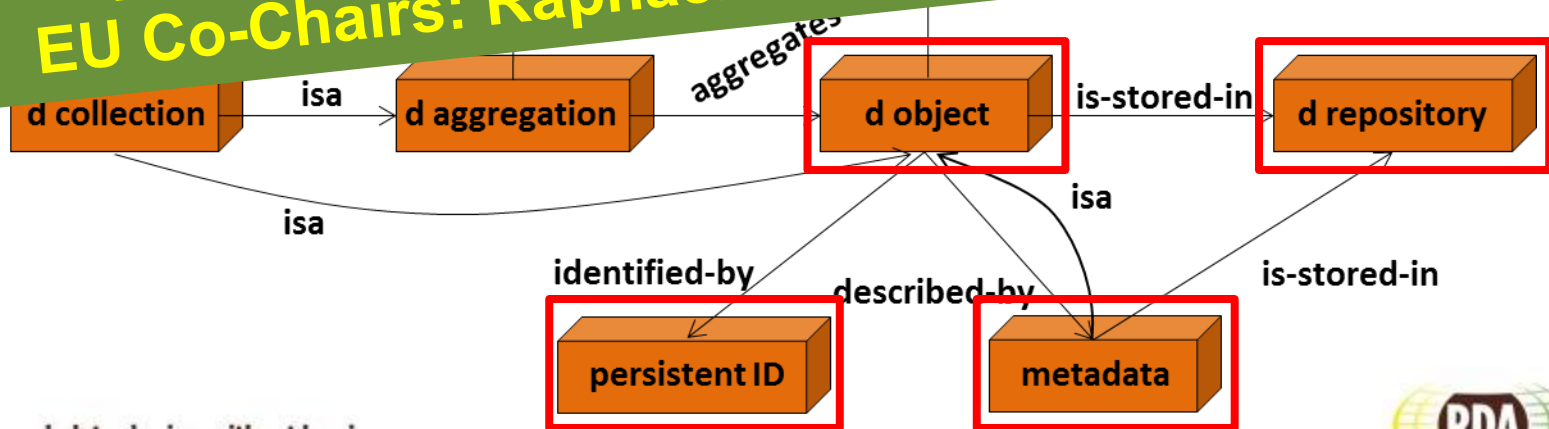
Welcher Ansatz?

- wie soll man die vielen technischen & sozialen Barrieren überwinden, um in ca. 5 Jahren fit zu sein?
- RDA wählt den Weg des Internet:
 - **bottom-up** und **rough consensus** als Basis Prinzipien
 - Engagement der **Daten-Experten, Daten-Wissenschaftler**, etc.
 - alles andere als ein **globaler Ansatz** wäre unsinnig
 - RDA entwickelt **keine Standards** – kostet zu viel Zeit – aber RFCs
 - offizieller Start im **März 2013** in Göteborg
 - alles relativ neu – noch keine **RDA Kultur**
 - **erste Resultate** von 4 WGs (siehe **Flyer**)
 - neue vielversprechende Aktivitäten
 - viel Engagement mehrerer exzellenter Leute (gerade WG Chairs Meeting in Washington, EU muss vorne dabei bleiben)



WG Data Foundation & Terminologie

- Resultat: ein **simples Basis-Modell** und eine **Basis RDA Terminologie** (nicht als fertiges Produkt zu sehen)
- viel an
- **PIDs stehen am Beginn der Vertrauenskette und wir wissen jetzt wie es geht bzw. was wichtig ist**
- **ähnliche Bedeutung wie TCP/IP für das Internet**
- **Metadaten sind essentiell in der anonymen Datenwelt**
- **nehmen Sie Teil über das Semantic Wiki TeD_T**
http://smw-rda.esc.rzg.mpg.de/index.php/Main_Page
- **EU Co-Chairs: Raphael Ritz, Peter Wittenburg (MPG)**



WG Data Type Registry

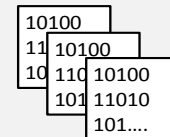
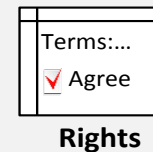
- Resultat: ein Registry für Datentypen
- man bekommt eine unbekannte Datei, zieht es auf die DTR und der Inhalt wird sichtbar gemacht
- erweitertes MIME Type Konzept
- no free lunch: jemand



- das NIST hat bereits mit einigen Communities weitgehende Projekte ausgearbeitet
- in EU müssen wir ebenfalls das Potential erkunden
- **EU-Co-Chair: Daan Broeder (MPG)**

- der Code ist Anfang 2015 verfügbar
- NIST/EUDAT werden erste DTR Services anbieten
- PIT Demo operiert bereits zusammen mit DTR

Domain of Services



WG PID Information Types

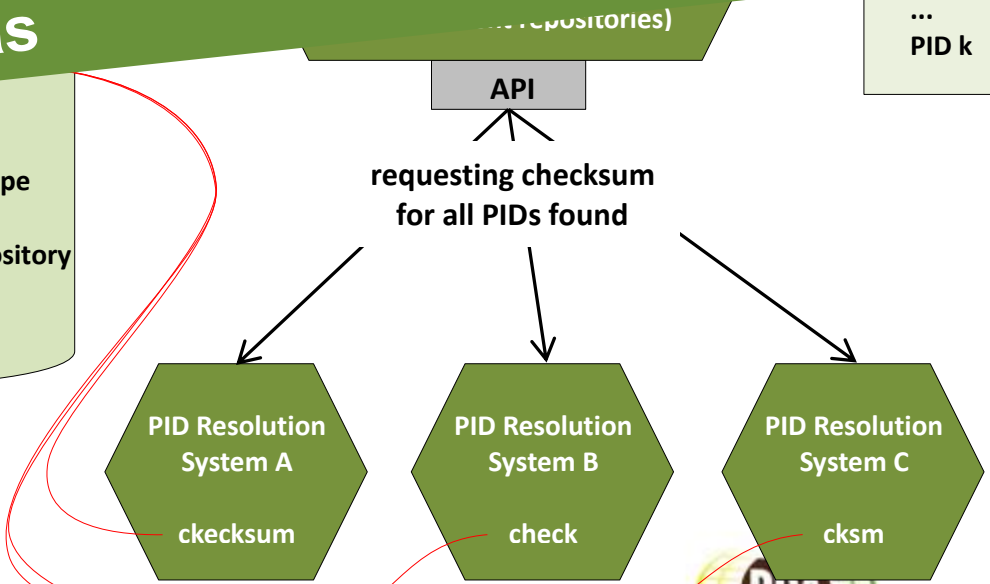
- Resultat: ein generisches API und ein Satz von Basis Attributen
- ein PID Record ist wie ein Pass (Nummer, Photo, Exp-Datum, etc.)
- es wird mehrere PID Service-Provider geben

das Umgehen mit PID Records wird vereinfacht
 weltweite Interoperabilität wird erzeugt
 EU Co-Chair: Tobias Weigel (DKRZ)
 mehr später von Tobias

in Operation
 zusammen mit
 DTR

CKSM	checksum
CKSM_T	checksum type
RoR	owning repository
MD	nath to MD

defined in DTR



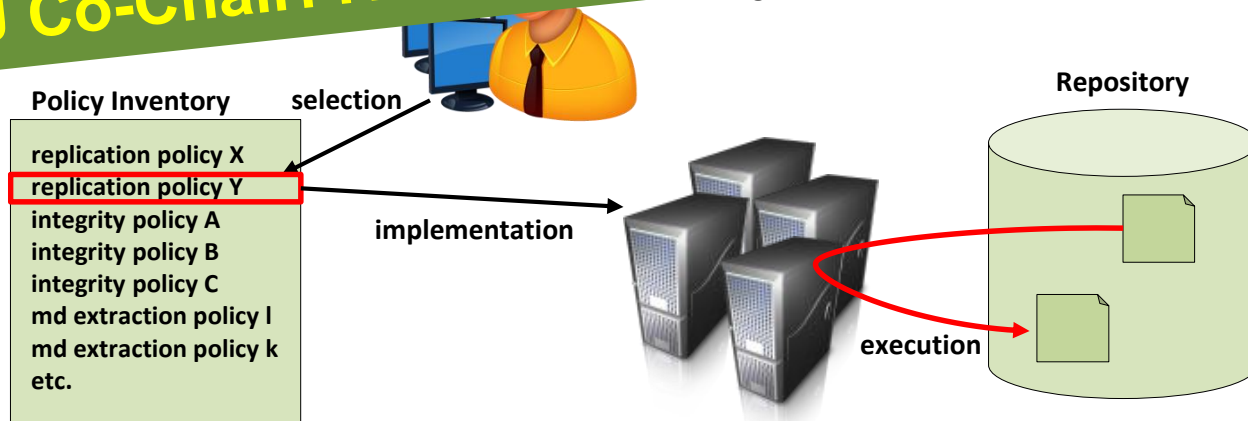
WG Practical Policies

- durch Krankheitsfall noch nicht ganz fertig

- Practical Policies = ausführbare Workflow Anwendung

- Resultat zum D5

- erhebliche Vereinfachung für Datenmanager
- Ermöglichung von Qualitäts-Prüfungen
- erhebliche Verbesserung des Vertrauens
- ein Baustein auf dem Weg zu reproduzierbarer Datenwissenschaft
- **EU Co-Chair: Rainer Stotzka (KIT)**

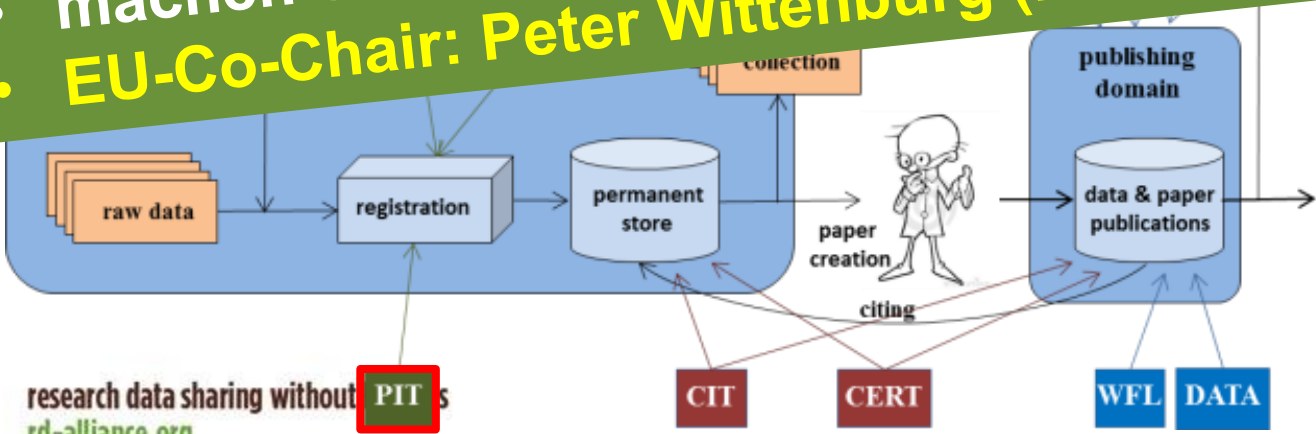


Neue Aktivitäten

- die ersten WGs haben sich “soloistisch” urgente Probleme vorgenommen
- bereits seit längerem intensive Interaktion zwischen Chairs
- nunmehr Gründung einer IC

• sehr engagierte Diskussionen
 • welches sind Komponenten und Services, Interfaces
 • erstes Resultat: brauchen eine akzeptierte Spezifikation für eine Registry von Trusted Repositories (-> EUDAT, EGI, etc.)
 • machen Sie mit – einfach registrieren über Web-Seite
 • **EU-Co-Chair: Peter Wittenburg (MPG)**

auch Fokus im Bereich Daten-Publishing



Übersicht

Agriculture Data Interest Group (IGAD)	IG	com
Biodiversity Data Integration	IG	com
Wheat Data Interoperability	WG	com
ELIXIR Bridging Force	IG	com
Metabolomics	IG	com
Structural Biology	IG	com
The BioSharing Registry: connecting data	WG	com
Toxicogenomics Interoperability	IG	com
Digital Practices in History and Ethnography	IG	com
Data for Development	IG	com
Urban Quality of Life Indicators	WG	com
Geospatial	IG	com
Marine Data Harmonization	IG	com
RDA/CODATA Materials Data, Infrastructure & I	IG	com
Research data needs of the Photon and Neutron	IG	com

Data Citation	WG	fab
Data Fabric	IG	fab
Data Foundation and Terminology	WG	fab
Federated Identity Management	IG	fab
Reproducibility	IG	fab
Data Type Registries	WG	fab
Long tail of research data	IG	fab
Data Description Registry Interoperability	WG	fab
Data in Context	IG	fab
Metadata	IG	fab
Metadata Standards Directory	WG	fab
Research Data Provenance	IG	fab
PID Information Types	WG	fab
PID	IG	fab
Domain Repositories I	IG	fab
Big Data Analytics	IG	fab
Brokering Governance	WG	fab
Brokering	IG	fab
Practical Policy	WG	fab
Service Management	IG	fab

Übersicht

Ethics and Social Aspects of Data	IG	gen	RDA/WDS Certification of Digital Repositories	IG	gen
RDA/CODATA Legal Interoperability	IG	gen	Repository Audit and Certification DSA-WDS	WG	gen
RDA/WDS Publishing Data Bibliometrics	WG	pub	Libraries for Research Data		gen
RDA/WDS Publishing Data Cost Recovery for Data	IG	pub	Active Data		gen
RDA/WDS Publishing Data					gen
RDA/WDS Publishing Data					gen
RDA/WDS Publishing Data					gen
RDA/WDS Publishing Data					gen
Archiving of interactive /dynamic data	IG	arch	Education and Training on handling of research data	IG	gen
Preservation e-Infrastructure	IG	arch	Summer Schools in Data Science and Cloud	WG	gen
			Standardisation of Data Categories and Codes	WG	gen

TAB ist dabei, eine Art Clustering einzuführen
übersichtlicher für Außenwelt
einfacher, um Überlappungen festzustellen

Zusammenfassung

- erste WGs waren die **Versuchskaninchen** – mussten Prozeduren als Teil der RDA-Kultur ausarbeiten
- selbst erstaunt, dass nach 18 Monaten erste **belastbare Ergebnisse** vorliegen
- zumeist **unbezahltes Engagement** als Basis
- 1. Phase ist vorbei
- **2. Phase beginnt**
 - Einordnung aller Aktivitäten in umfangreichere Landschaft
 - gezieltes Starten neuer Aktivitäten aus bottom-up Diskussionen (ml Registratur von Repositorien für Data Fabric, etc.)
 - Organisieren der Übernahme von Resultaten
 - Hilfen und Support organisieren

Nehmen Sie Teil am Brückenbauen

16

Nächstes Plenary P5:
Plenary P6:

9-11. März San Diego
September Paris



RDA is about building the social and technical bridges that enable global open sharing of data.

Researchers, scientists, data practitioners from around the world are invited to work together to achieve the vision

Dank für Ihre Aufmerksamkeit



<http://rd-alliance.org>

<http://europe.rd-alliance.org>