



Das DFG Verbundprojekt IBF – Datenfluss und Etablierung eines institutionellen Datenzentrums

A. Kehl

mit S. Jablonski, D. Neubacher, P. Poschlod,
G. Rambold, T. Schneider, D. Triebel, B. Volz, M. Weiss

Contents

- Demands of Science policy
- Traditional and aspired dataflow
- IBF-Project
 - dataflow
 - Institutional Data Repositories
- Conclusions

Demands of Science Policies

- DFG (1998):
„Vorschläge zur Sicherung guter wissenschaftlicher Praxis“:
Storage for a minimum of ten years (mostly local databases)
- Allianz-Partnerorganisationen (2008):
„Schwerpunktinitiative Digitale Information“: Long-term storage
and guarantee of permanent availability and quality control in
“Data Repositories”

Traditional Dataflow

Field

Gathering of Primary Data

Pen & Paper

Project-sided Computer (local)

Data Storage and Analysis

Data Digitalization

- Post-processing of data
- Further analysis (e. g., statistics)

Data items

Genotypical and phenotypical distinction of *Salix* × *rubens* 175

Table 5. Kruskal-Wallis ANOVA: median of phenotypical traits between the four ITS-genotypes

	H	P	ITS-genotype			
			1	2	3	4
Leaf length (mm)	65.64	<0.001	79.15 ^a	69.15 ^{bc}	72.75 ^b	57.90 ^f
Leaf width (mm)	142.42	<0.001	1530 ^a	12.15 ^b	12.00 ^b	11.00 ^b
Leaf length/width	101.22	<0.001	4.95 ^a	5.68 ^{bc}	6.00 ^{bc}	5.44 ^c
Petiole length (mm)	98.15	<0.001	9.90 ^a	7.75 ^b	7.55 ^b	4.40 ^f
Pubesence upper leaf surface	167.18	<0.001	0.08 ^a	2.74 ^{bc}	2.66 ^{bc}	15.08 ^d
Pubesence lower leaf surface	174.04	<0.001	0.12 ^a	0.96 ^{bc}	7.88 ^d	23.04 ^e
Dry weight of leaf (g)	137.10	<0.001	0.49 ^a	0.29 ^b	0.30 ^b	0.22 ^c
Leaf water content (g)	98.68	<0.001	0.87 ^a	0.61 ^b	0.63 ^b	0.47 ^c
Specific leaf weight (mg/cm ²)	36.34	<0.001	48.29 ^a	46.14 ^{ab}	46.49 ^{ab}	41.87 ^b
Height growth (cm)	56.43	<0.001	32.33 ^a	46.19 ^b	46.33 ^b	41.33 ^b
Diameter growth (mm)	17.51	<0.001	4.60 ^a	8.55 ^{bc}	5.90 ^b	4.73 ^c
Clonal form (shoots/height)	21.47	<0.001	0.94 ^a	1.96 ^{bc}	1.23 ^b	1.08 ^{bc}
Foliation phenology 2003 (days after April 1st)	101.90	<0.001	16 ^a	12 ^b	14 ^b	22 ^c
Foliation phenology 2004 (days after April 1st)	119.09	<0.001	16 ^a	3 ^b	3 ^b	24 ^c
Foliation phenology 2005 (days after April 1st)	133.93	<0.001	14 ^a	2 ^b	5 ^b	16 ^c

H-test statistics, P significance level. Highest value per variable are in bold, different letters indicate significant differences

the investigated clones with reliable determined herbarium material and literature descriptions led to the affiliation to three taxa: *S. alba* L. (including clones 6, 9, 10, 17, 18), *S. fragilis* L. (including clones 1, 2, 3, 4, 7, 8, 11, 13, 15, 16), and their hybrid, *Salix* × *rubens* Schrank (including clones 5, 12, 14, 19). Both last mentioned taxa are used in the sense of Christensen and Jonell (2005). The fact that the ITS-sequences of ITS-genotype 1 and 4 were completely conform with the already published sequences of *S. fragilis* and *S. alba* from the work of Leskinen and Alstrom-Rapaport (1999), and furthermore, that the ITS-genotypes 2 and 3 revealed additivities on variable positions between *S. fragilis* and *S. alba*, which supports the hybrid origin of these four clones, validated this affiliation.

Journal publication

STO 1	STO 2	ID Ind.	Höhe cm	d mm	Datum
A	1	2/56			
A	2	4/12	54	6,5	23.9.05
A	3	2/45	125	18,4	
A	4	2/11	192	32,1	
A	5	2/1	280	57,3	
A	6	2/50	105	14,2	
A	7	2/36	118	28	
A	8	4/43	139	17,9	
A	9	2/34	249	22	
A	10	4/1	166	30,5	
A	11	2/8			
A	12	4/2	186	19,5	
A	13	2/57	133,5	19,6	
A	14	2/55	123	24,2	
A	15	2/51	140	17,7	
A	16	2/35	178	22,6	
A	17	2/58	118	18,5	
A	18	2/53	166	24,2	
A	19	2/52	140	21	
A	20	2/56	112	16,2	
A	21	4/12			
A	22	2/45	74	12,7	
A	23	2/11	113	17,8	
A	24	2/1	187	28,1	
A	26	2/36	84	20,4	
B	1	4/13			
B	2	2/34			26.9.
B	3	4/1	179	20	

Microsoft Excel - Faktorenanalyse - AKTUELL.xls

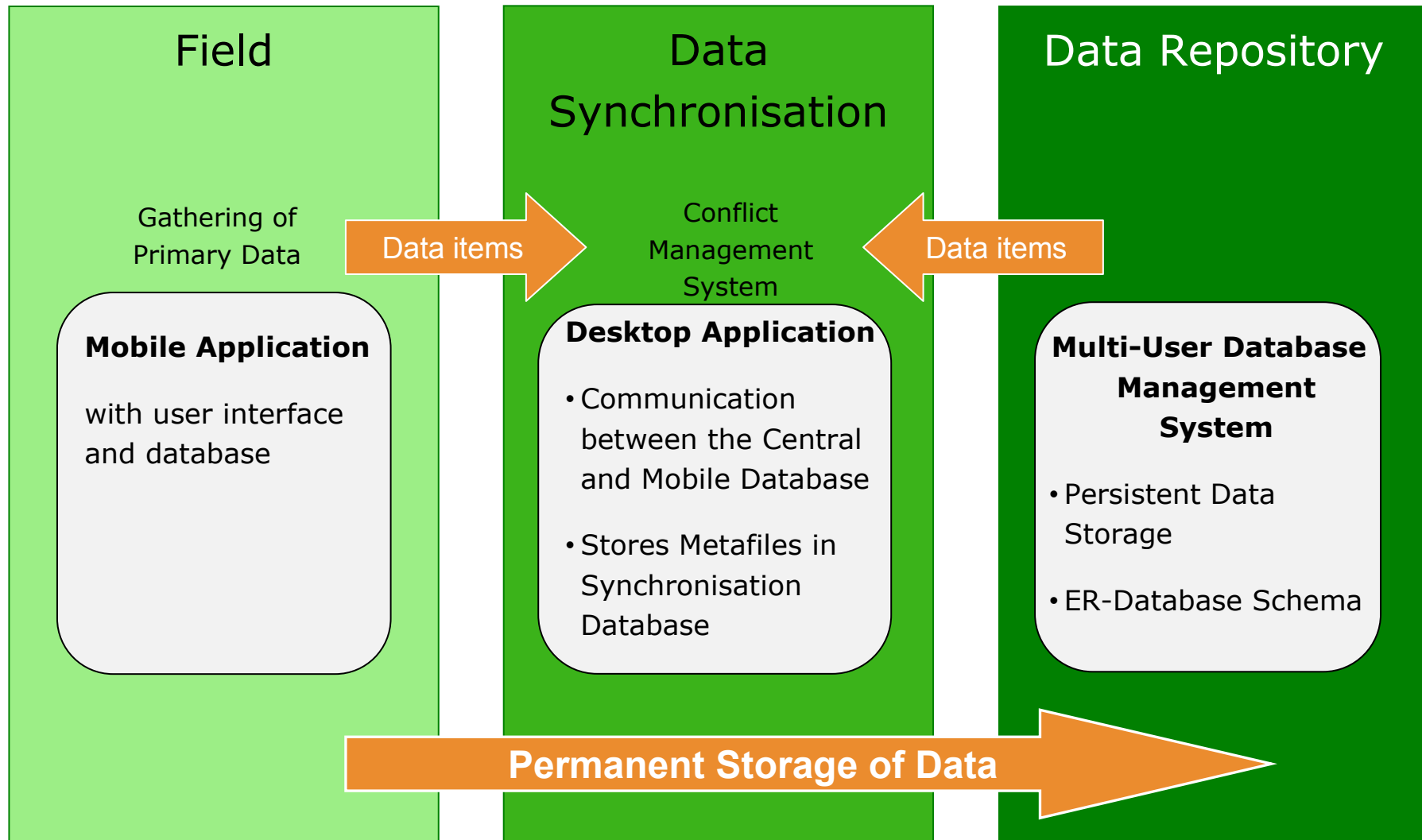
Datum: Bearbeiten Ansicht Einfügen Format Extras Daten Ersteller: 2 Adgbe PCF

Frage hier eingeben

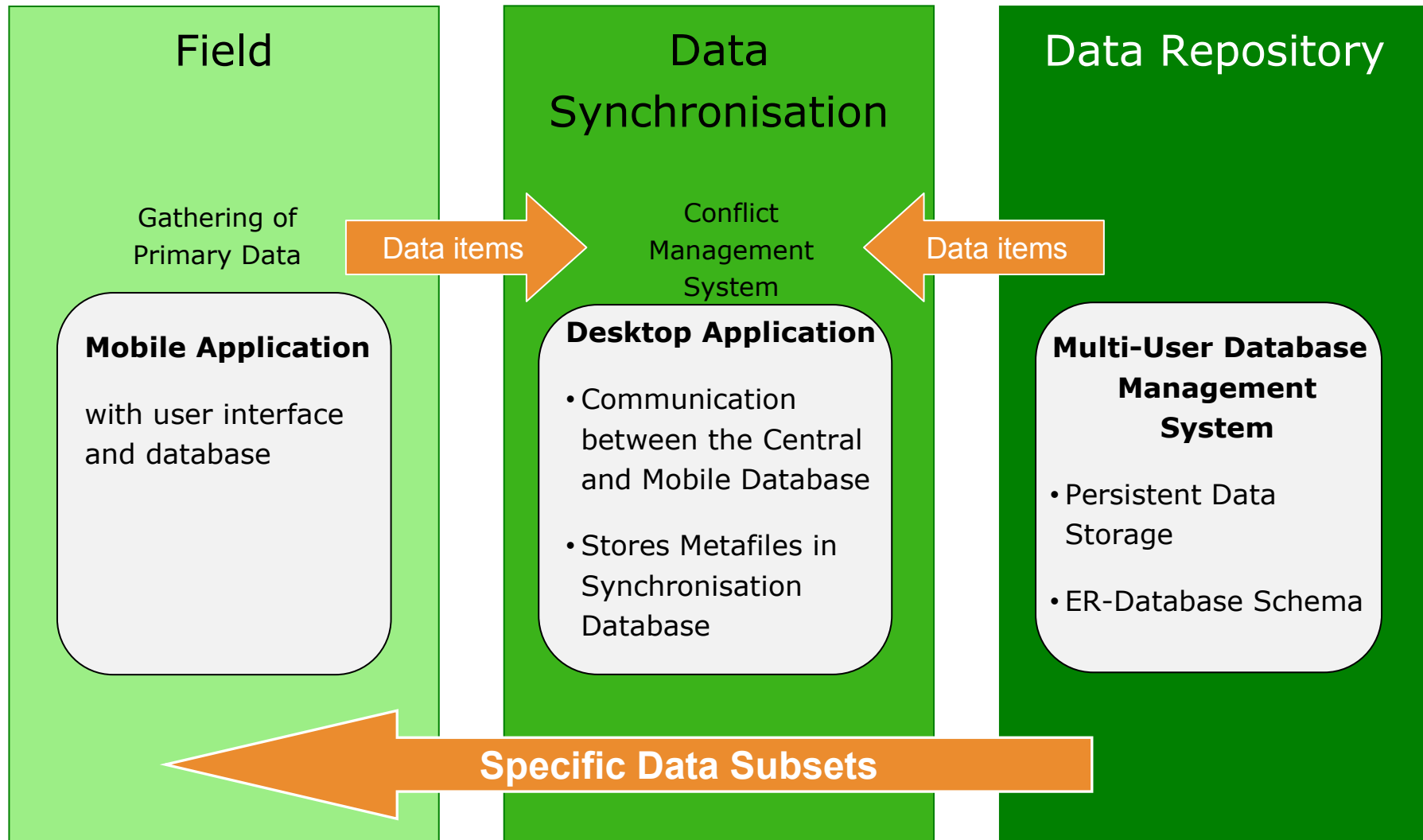
Q44

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2	2810	2890	2910	2930	2950	2970	2990	3010	3030	3050	3070	3090	3110	3130	3150	3170	3190	3210	3230	3250	3270
3	STO	STO	ID	Inv	Positiv	ID	Inv	Tax	Col												
4	A	5	2/1	1	1	1	1	2	0	1	86	16,8	0,2	0,28	0,591	1,157	66,18959	12,70343	9,304575	4,5653	
5	A	24	2/1	1	1	1	1	2	0	1	77,6	18,3	0,12	0,16	0,46	0,822	64,39226	11,19177	9,24374	4,2195	
6	B	17	2/1	1	1	1	1	2	0	2	81	17,6	0,04	0,08	0,491	0,918	65,15259	11,20114	8,769683	4,5803	
7	C	10	2/1	1	1	1	1	1	2	2	98,2	20,4	0,16	0,04	0,628	1,291	67,27462	15,74006	7,979641	4,8933	
8	D	5	2/1	1	1	1	1	2	0	1	92,9	20,2	0,24	0,49	0,714	1,375	65,62097	15,72869	9,69306	4,5991	
9	E	2	2/1	1	1	1	1	1	0	1	88	18,9	0,2	0,95	0,6013	1,026	79,63398	13,068	9,201867	4,6691	
10	E	21	2/1	1	1	1	1	2	1	1	84,7	17,6	1,28	2,88	0,607	1,102	64,48216	11,7128	10,36473	4,8324	
11	F	16	2/1	1	1	1	1	2	0	1	86	17	0,18	0,2	0,599	1,16	65,94656	11,48714	9,205	3,3889	
12	G	19	2/1	1	1	1	1	2	0	1	85,2	17,5	0,36	0,76	0,536	1,03	65,77257	11	9,205	3,2	4,9173
13	H	5	2/1	2	1	1	1	2	0	1	87,5	19,3	0,04	0,04	0,98	1,213	67,65198	13,26	9,205	3,2	4,5403
14	H	29	2/1	2	1	1	1	2	0	1	104,7	20,1	0,16	0,16	0,742	1,622	68,67252	16,6351	9,205	3,2	4,5928
15	I	11	2/1	2	1	1	1	2	1	1	94,2	18,9	0,24	0,04	0,692	1,214	63,6936	13,9887	9,205	3,2	4,5928
16	J	15	2/1	2	1	1	1	2	0	1	91,8	18,9	0,08	0,16	0,611	1,272	67,55178	13,6323	9,205	3,2	4,5928
17	K	5	2/1	2	1	1	1	2	0	1	85,6	19,4	0,2	0,04	0,952	1,114	65,68396	12,74303	9,205	3,2	4,5928
18	K	22	2/1	2	1	1	1	2	0	1	109,2	20,7	0,08	0,16	0,719	1,615	69,19402	16,6221	9,205	3,2	4,5928
19	A	4	2/11	1	2	2	3	2	2	2	4	65	13,1	3	11,4	0,28	0,539	65,81197	6,690357	8,370256	4,2
20	A	23	2/11	1	2	2	2	2	2	2	4	75,1	13,3	2,2	10,8	0,352	0,69	66,21881	7,84795	8,970495	5,6951
21	B	16	2/11	1	2	2	2	2	2	2	4	76,2	11,9	3,52	67,2072	0,691	67,2072	6,676686	9,18178	6,6701	
22	C	9	2/11	1	2	2	3	2	2	2	4	67,3	12,4	2	9,32	0,334	0,586	63,69565	6,556843	10,18767	5,4168
23	D	5	2/11	1	2	2	2	2	2	2	2	71	13,7	5,88	13,72	0,378	0,684	64,5293	7,642643	9,639226	5,2315
24	E	1	2/11	1	2	2	2	2	2	2	2	75,3	12,6	2,48	6,44	0,326	0,7	65,2812	7,4547	6,74616	6,9947
25	E	20	2/11	1	2	2	2	2	2	2	2	66,6	10,3	2,12	8,96	0,287	0,403	62,96878	4,880557	10,34809	5,5184
26	F	15	2/11	1	2	2	2	2	2	2	2	85,6	10,9	3,36	11,32	0,28	0,575	67,25146	5,18171	9,967656	5,9247
27	G	12	2/11	1	2	2	2	2	2	2	2	76,2	10,8	1,52	4,44	0,27	0,833	70,09897	6,381257	6,822292	6,9576
28	H	4	2/11	2	2	2	2	2	2	2	2	63,3	12,4	5,52	12	0,285	0,4956	52,91711	6,167229	9,244201	5,1876
29	H	25	2/11	2	2	2	2	2	2	2	2	61,7	10,9	4,6	10,8	0,36	0,784	65,67669	6,284164	10,7401	5,6693

Aspired Dataflow



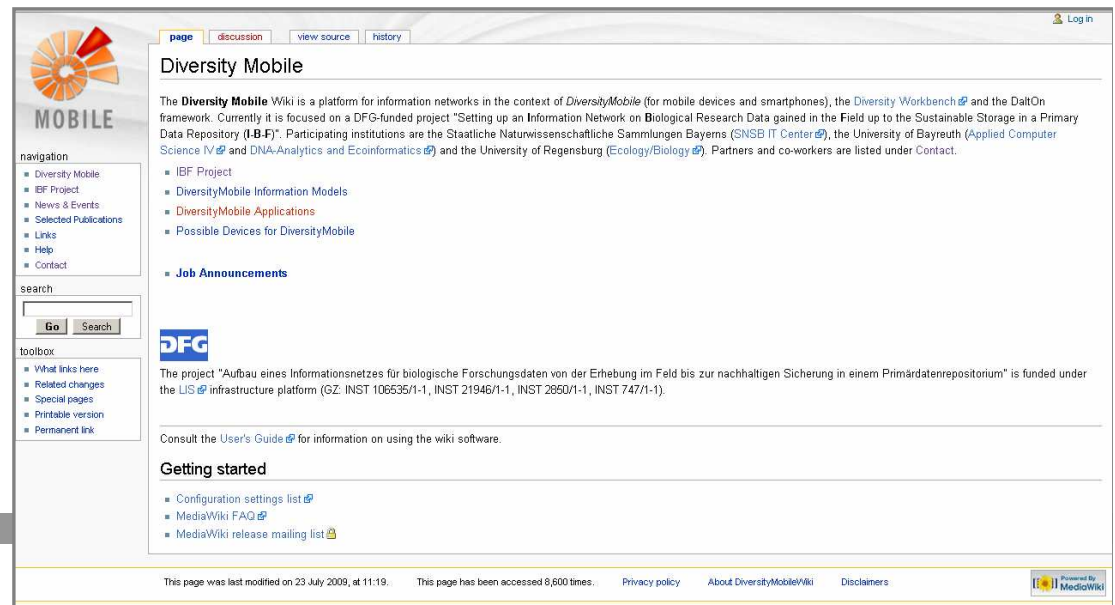
Aspired Dataflow



IBF Project

- Setting up an **I**nformation Network on **B**iological Research Data gained in the **F**ield up to the Sustainable Storage in a Primary Data Repository (**IBF**)

→ www.diversitymobile.net



The screenshot shows the Diversity Mobile Wiki page for the IBF Project. The page title is "Diversity Mobile". The main content area contains the following text: "The **Diversity Mobile** Wiki is a platform for information networks in the context of *DiversityMobile* (for mobile devices and smartphones), the *Diversity Workbench* and the *DatOn* framework. Currently it is focused on a DFG-funded project "Setting up an Information Network on Biological Research Data gained in the Field up to the Sustainable Storage in a Primary Data Repository (**IBF**)". Participating institutions are the Staatliche Naturwissenschaftliche Sammlungen Bayerns (SNSB IT Center), the University of Bayreuth (Applied Computer Science IV and DNA-Analytics and Ecoinformatics) and the University of Regensburg (Ecology/Biology). Partners and co-workers are listed under Contact."

The page also features a navigation menu on the left with links to "Diversity Mobile", "IBF Project", "News & Events", "Selected Publications", "Links", "Help", and "Contact". There is a search box and a toolbox with links to "What links here", "Related changes", "Special pages", "Printable version", and "Permanent link".

The DFG logo is prominently displayed, followed by the text: "The project "Aufbau eines Informationsnetzes für biologische Forschungsdaten von der Erhebung im Feld bis zur nachhaltigen Sicherung in einem Primärdatenrepositorium" is funded under the LIS infrastructure platform (GZ: INST 106535/1-1, INST 21946/1-1, INST 2850/1-1, INST 747/1-1)."

At the bottom of the page, there is a footer with the text: "This page was last modified on 23 July 2009, at 11:19. This page has been accessed 8,600 times. Privacy policy About DiversityMobile/Wiki Disclaimers".

IBF Project

Establishment of dataflow from and to an institutional data repository

- Data gathering in the field via mobile device (PDA, Smart Phone) with software (*DiversityMobile*) for
 - data entry
 - accessing taxonomy and ecology descriptor presets
- Data transfer to institutional data repository (via *DiversityWorkbench* database component *DiversityCollection*)
 - data storage
- Data redistribution to end-users by data transfer to PDAs, wrappers (ABCD schema) and diverse applications for data analysis and presentation
 - data analysis
 - data presentation
 - data re-usage

IBFProject

Field: Diversity Mobile

Tool for Gathering
Observation Data

Mobile Application

- Temporary Data Storage
- Local Windows CE Database

- Access to external data sources (e. g. standard list of taxon names)
- GPS functionality
- Collection of multimedia data (images, video, audio)
- Previously recorded data accessible in the field
- All entered data to be transferred to the data repository

IBF-Project



- Long-term data storage
- Access to data history
- Individual observations are referable (GUIDs, LSIDs)
- Selective data export for local post-processing
- Publication (traditional pathway, Internet)

Data Repository:
SNSB IT Center

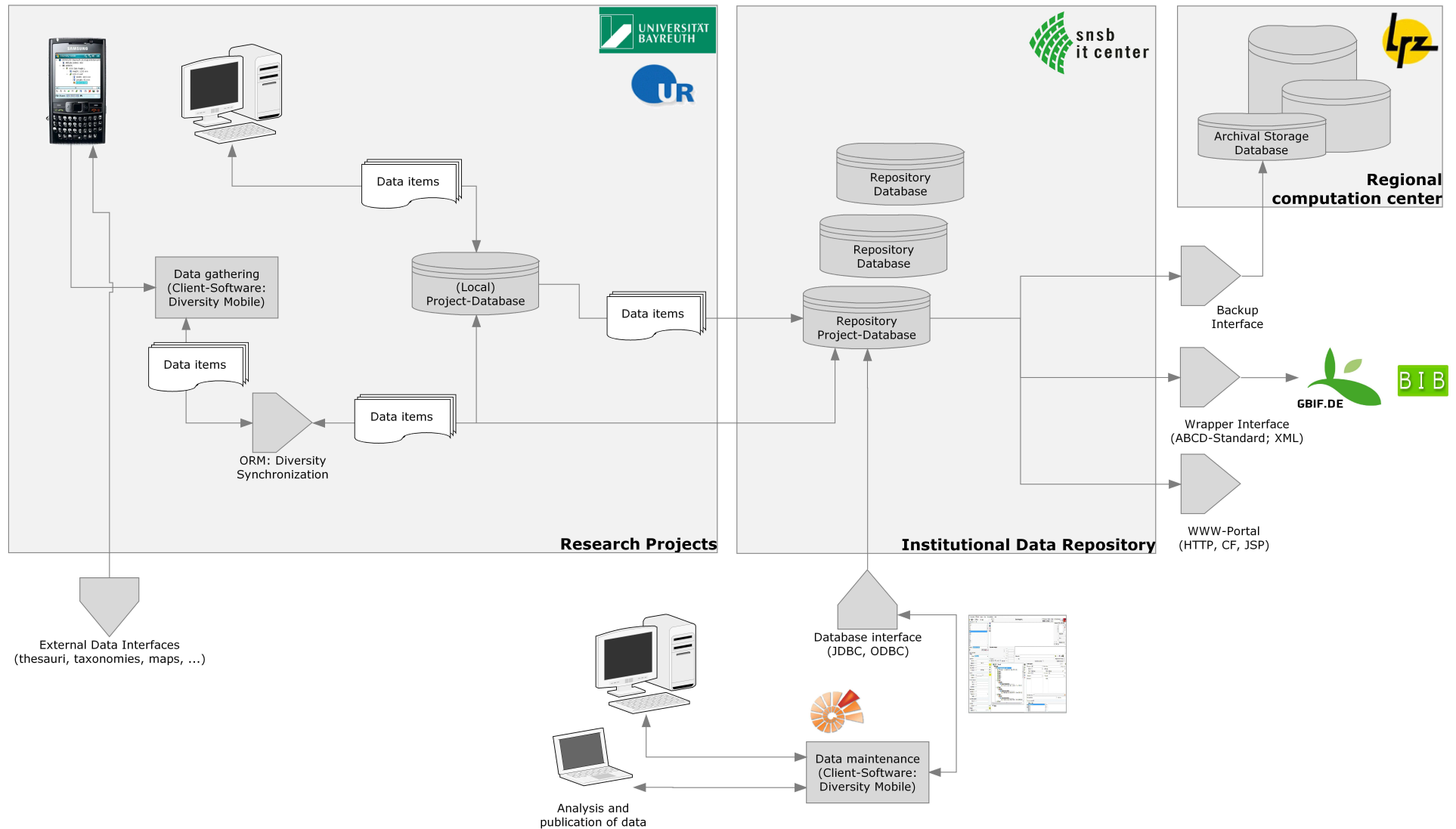
Data Storage
System: Diversity
Workbench

**Multi-User Database
Management
System**

- Defined ER-Database Schema

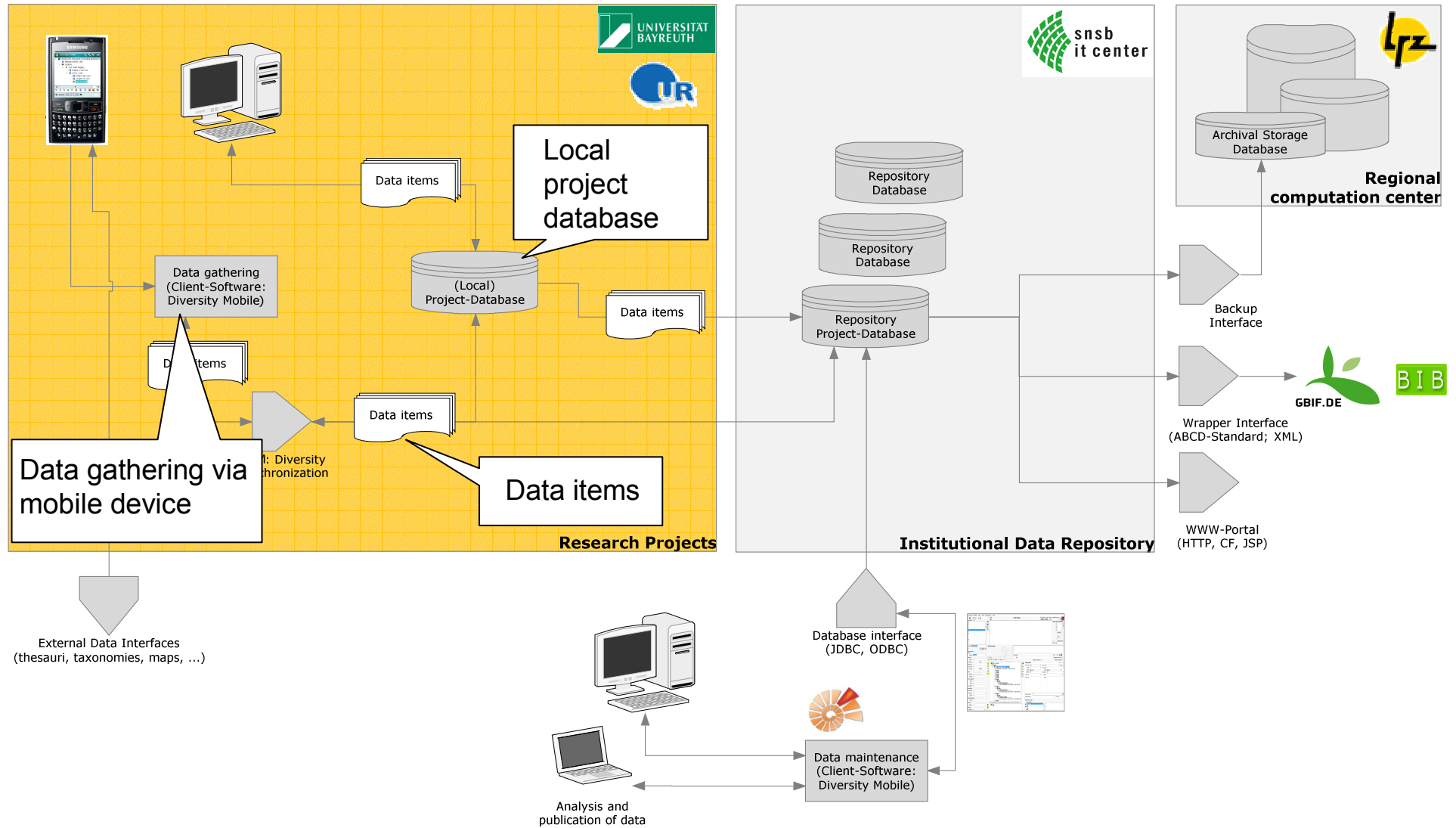
IBF Data Flow

Observatory Data, Data Flow



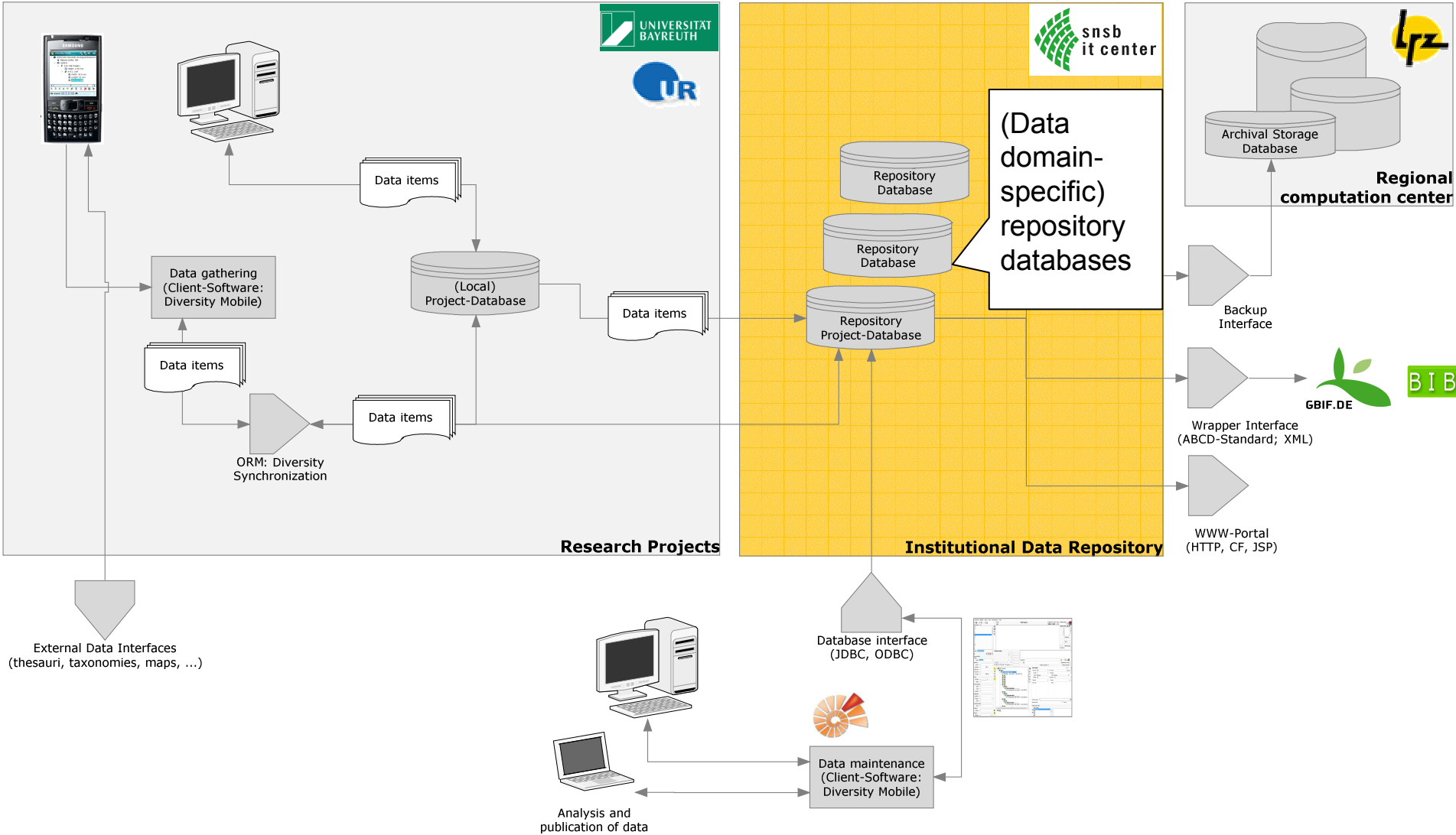
IBF Data Flow

Observatory Data, Data Flow



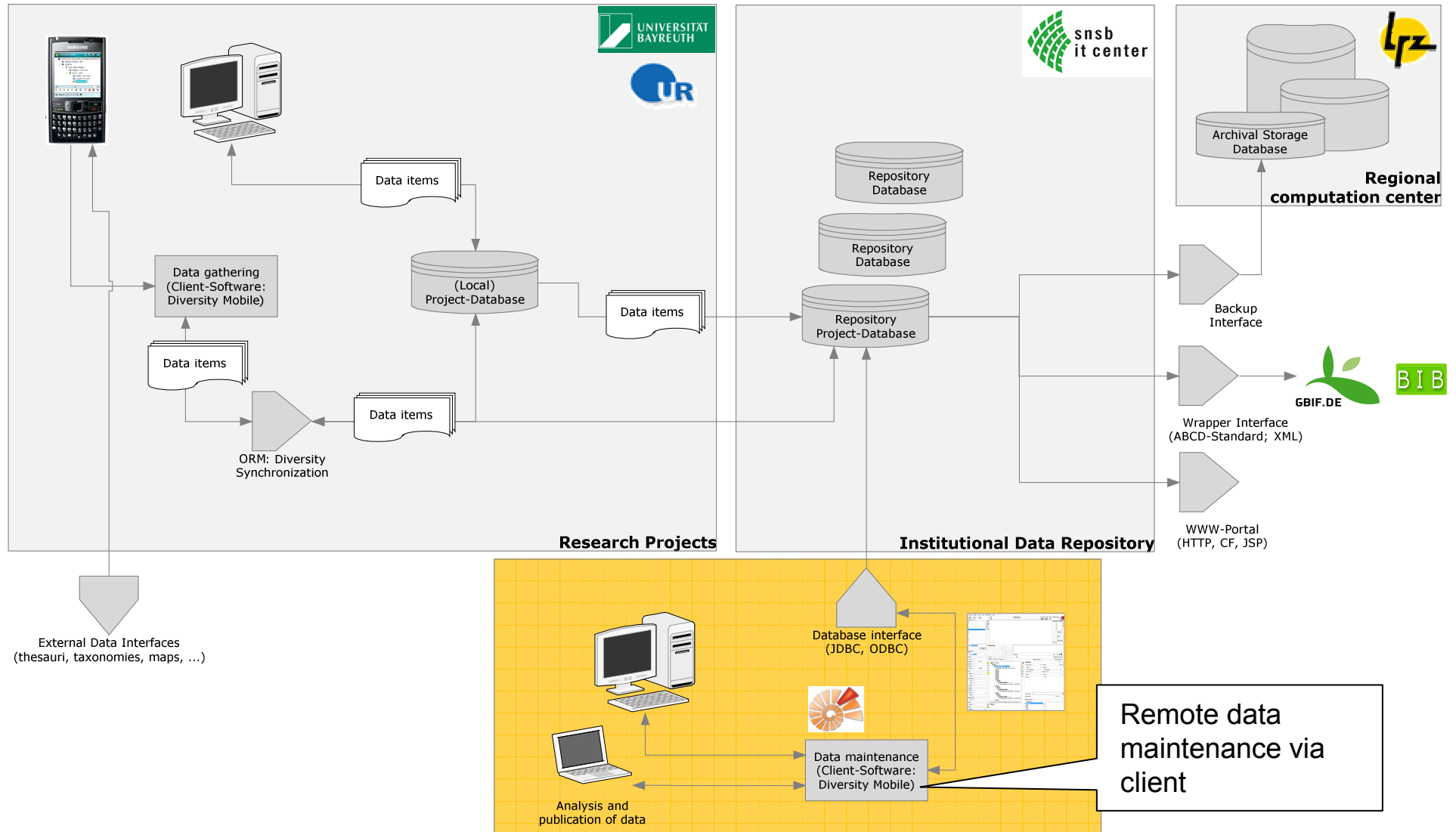
IBF Data Flow

Observatory Data, Data Flow



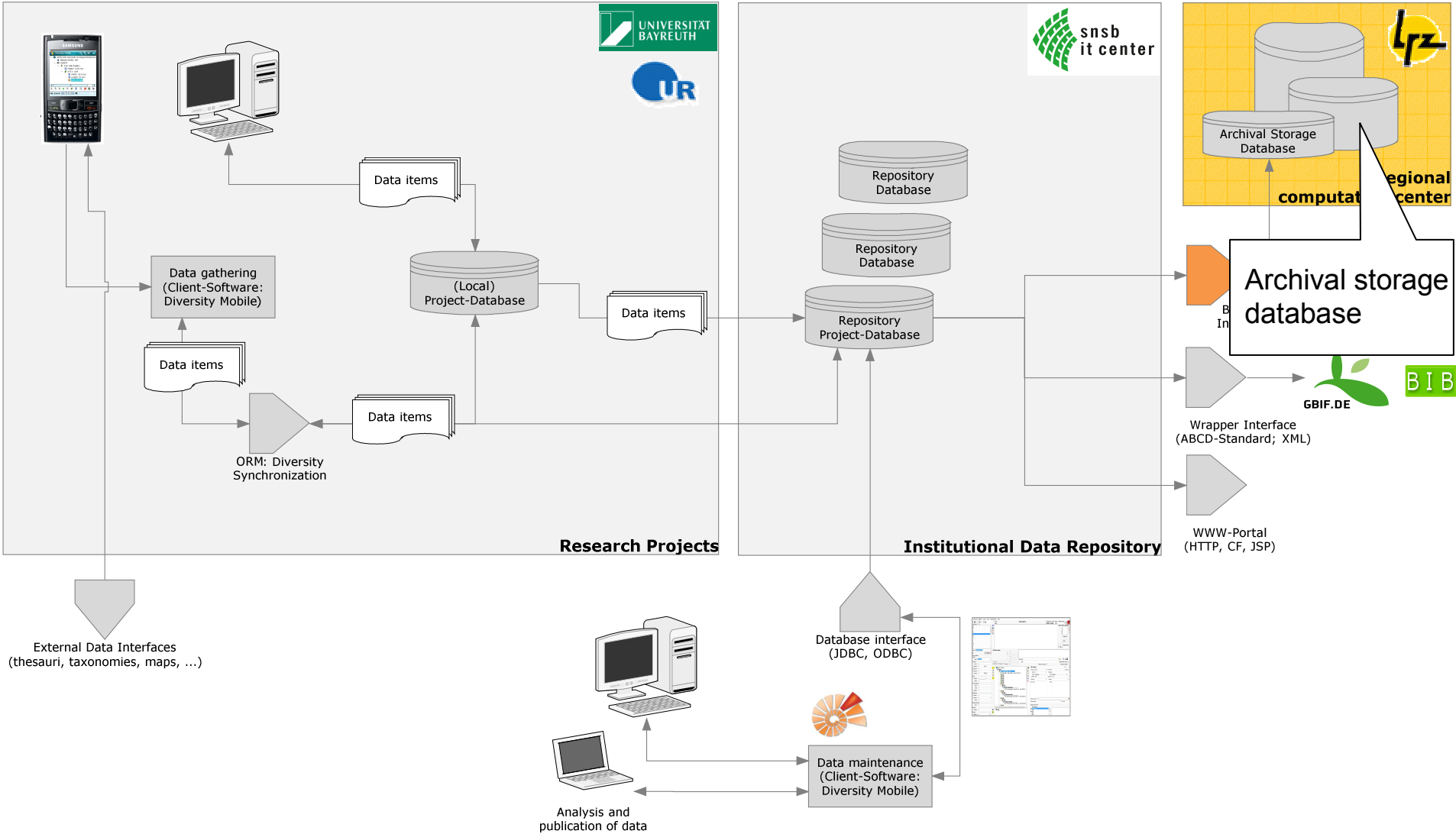
IBF Data Flow

Observatory Data, Data Flow



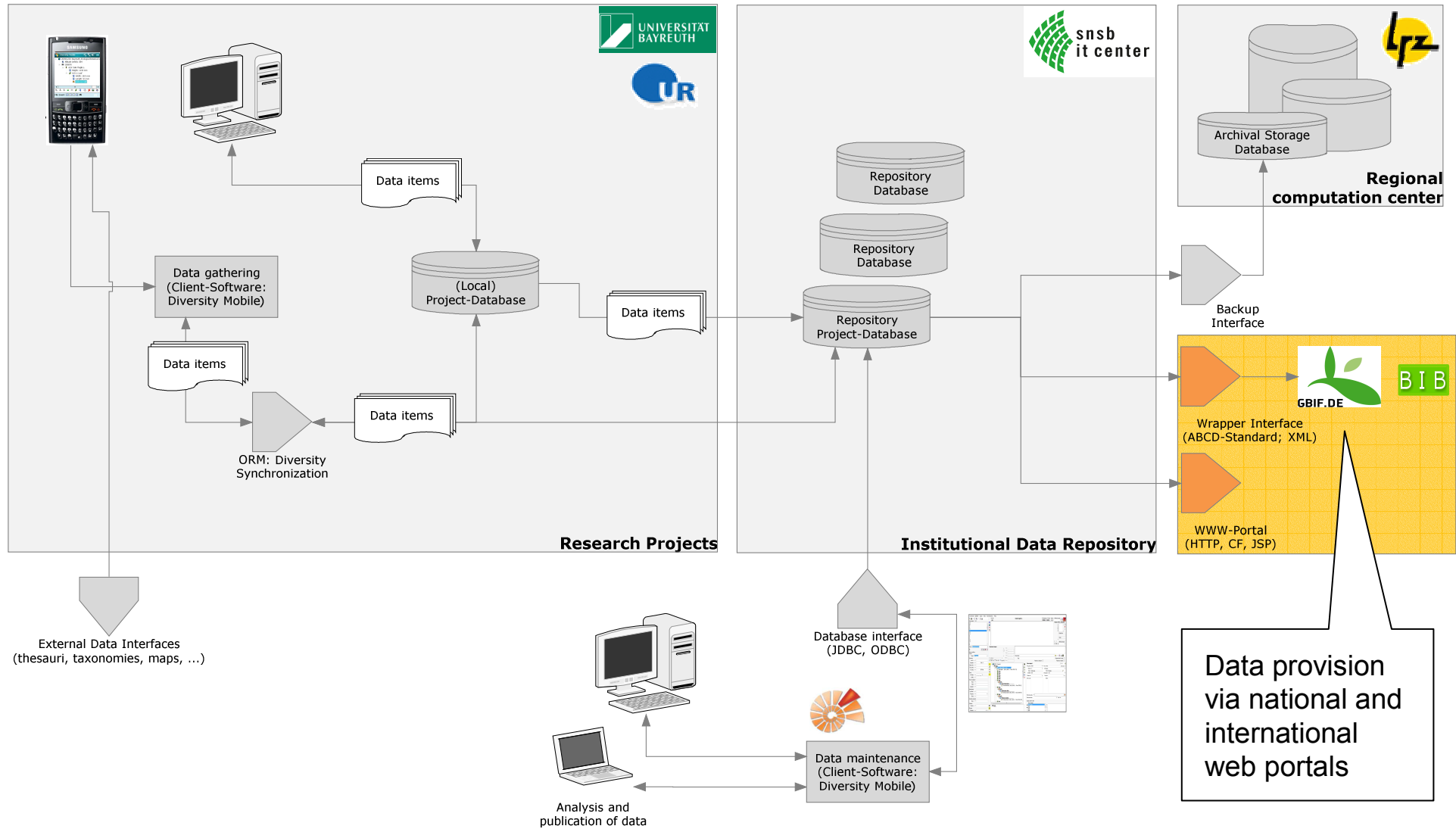
IBF Data Flow

Observatory Data, Data Flow



IBF Data Flow

Observatory Data, Data Flow

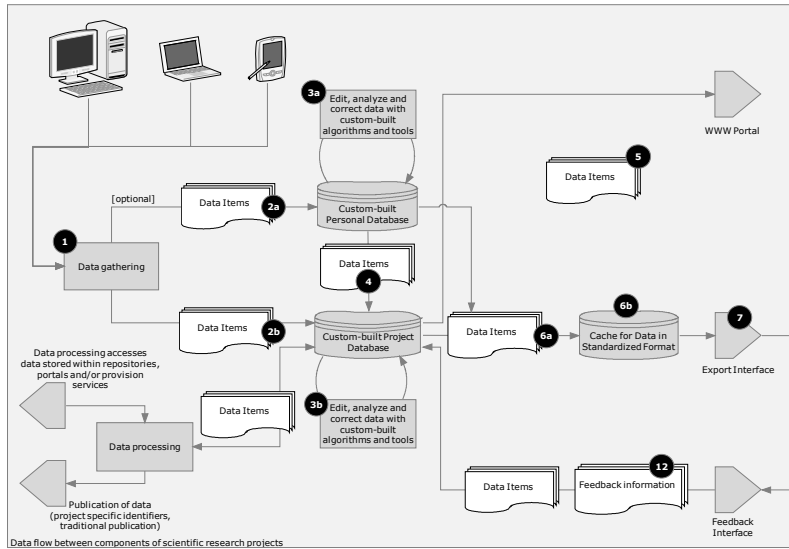


Data Repository Structure at SNSB

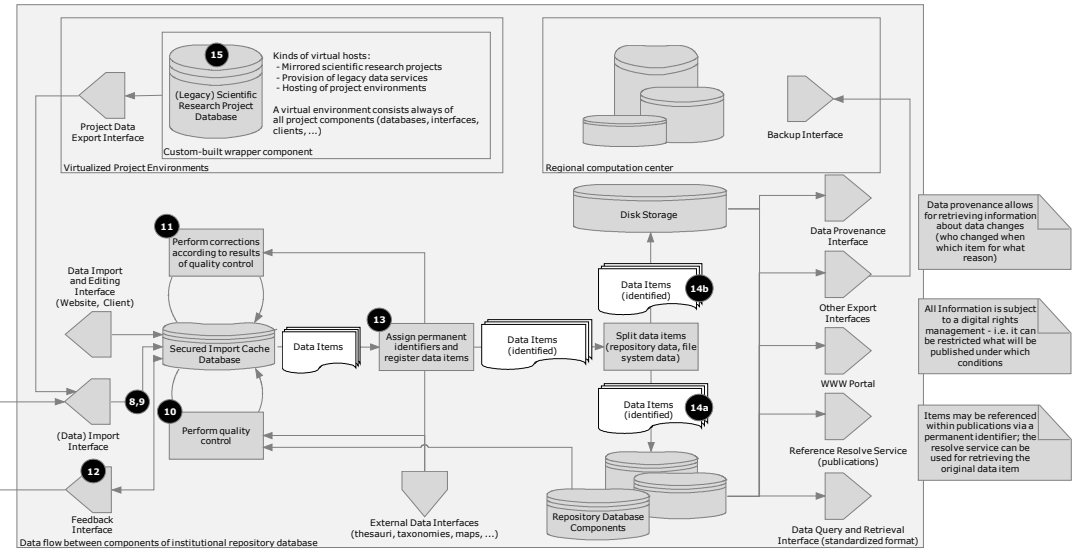
- Storage of structured research data
 - (Data domain-specific) structured 'standard data'
 - structured 'non-standard data'
- Provision/support of data standards (according to 'Biodiversity Information Standards', formerly TDWG)
 - Collection (and observation) data ABCD
 - Descriptive data SDD
 - Taxonomic data TCS
- Quality control and publication of primary data (→ persistent identifiers)

Data Repository Structure at SNSB

Research project

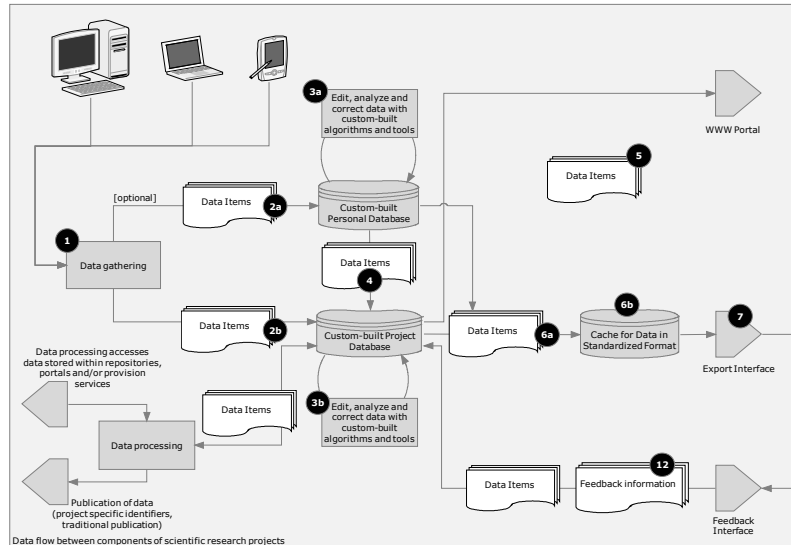


Institutional data repository

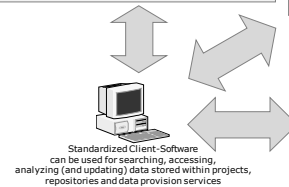
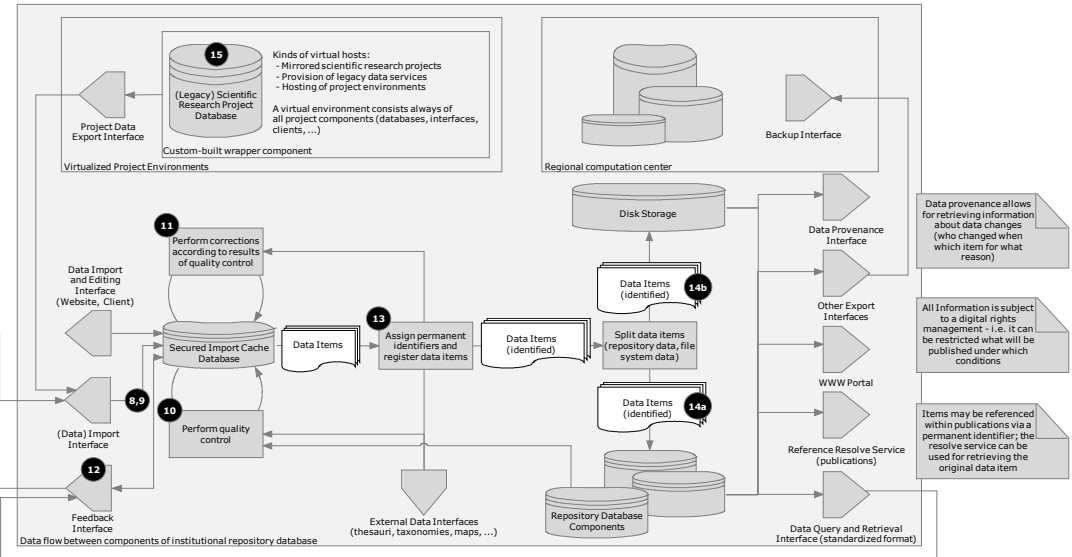


Data Repository Structure at SNSB

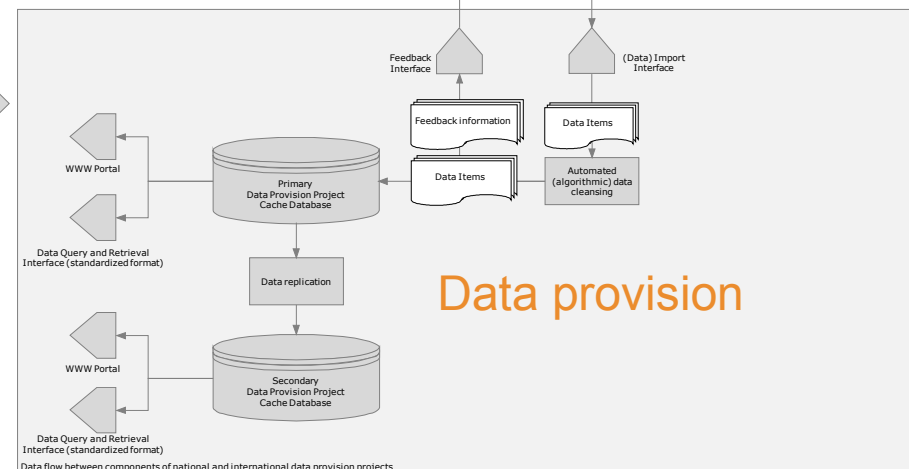
Research project



Institutional data repository



Data provision



Conclusions

Aspects of long-term storage of primary research data

- Scientific
 - Intersection and completion with standard data
 - Improvement of data quality
 - Precondition for reproducibility of data
- Economic
 - Access and subsequent use option on behalf of third parties
- Social
 - Credits to data owners and providers

Acknowledgements

Deutsche Forschungsgemeinschaft 

Förderbereich LIS – Informationsmanagement

(GZ: INST 106535/1-1, INST 21946/1-1, INST 2850/1-1, INST 747/1-1)