



Data Archiving in Energy Research

Iver Laueremann

Helmholtz-Zentrum Berlin für Materialien und Energie (HZB)

- **What is archiving?**
- **Why archiving of scientific data?**
- **Rules at HZB for archiving**
- **Recent history of archiving at HZB**
- **Discussion: what, who, when, where**
- **Data formats**
- **Obstacles to archiving in everyday work**
- **Examples for archiving at HZB**
 - Electronic lab-book
 - Scanning of paper lab-book
 - Storage of large amounts of data
- **Conclusions**

Archiving

Secure, long-time storage of scientific data and methods

- **secure: self-burned cds or dvds are not durable**
- **long-time: electronic data-formats change quickly and hardware fails with time, formats need to be adapted**

Requires meta-data, i.e. a description of the stored data that allows (expert) users the interpretation of the archived data

Makes later use of data by yourself or others possible!

- **usually only fraction of available data published**
- **new knowledge makes later use worthwhile**
- **for publishing of data, archiving is necessity condition**

Preservation of evidence for data from a publication

- **In many alleged cases of scientific misconduct original data had disappeared**

- **Data from space explorer Pioneer were stored in 4 different formats in 1979, however, in 1994 no hardware was available at NASA to retrieve the data from any of them.¹**
- **Data from 30 years of space exploration were stored on 1.2 million magnetic tapes in 1976 but in the mid-nineties many were useless due to insufficient description on the tapes.²**

*¹Hilmar Schmundt: Im Dschungel der Formate. In: Der Spiegel 26/2000. URL:
<http://www.spiegel.de/druckversion/0,1588,82510,00.html>*

*²Archimedes. Wir verlieren unser Gedächtnis, vom 04.05.1999. URL:
<http://www.artetv.com/hebdo/archimed/19990504/dtext/sujet1.html>*

- **Initiated by Wolfgang Fritsch (HMI library), 2 working groups (structure of matter department and energy department) on archiving established early in 2007**
- **In 2007/2008 discussed current situation and wrote 2 papers with recommendations**
- **July 2008: 2 day Workshop on archiving at HMI with participation of BESSY and external speakers**
- **Final paper**
- **December 2010: announcement of archive server by IT department**

- **What:** initially only data, on which a publication is based, has to be archived but in the long run all research data should be archived
- **Who:** head of organizational unit (i.e. institute director) is responsible but each scientist should be educated and be responsible for archiving
- **When:** on a regular basis; after a publication has appeared
- **Where:** central storage and support by IT is necessary
- **How:** As easy for users as possible, as flexible as possible, including current methods

THE RULES FOR ARCHIVING SCIENTIFIC DATA AT HZB

„Any scientific (raw) data on which some publication is based is documented and securely stored, for at least 10 years, at the organisational (department, institute) unit of its origin“

(from „Regeln zur Sicherung guter wissenschaftlicher Praxis und zum Verfahren bei wissenschaftlichem Fehlverhalten 14.06.2002.“

Based on the „Proposals for Safeguarding Good Scientific Practice“, January, 1998 by the Deutsche Forschungsgemeinschaft (DFG)

Where?

- Since Dec. 2010 „official“ archive server at HZB
- Data are stored on magnetic tape, can not be accessed unless removed from archive
- Path \\home\NB-Archiv\dauer\ordnername can be addressed by every internal HZB computer and filled with data
- Content of folders is copied on several magnetic tapes when no more changes detected, removed from hard drive
- Archiving over a period of 5 or 10 years
- Archives are checked once a year to ensure legibility
- Only members of respective department can access data

*(from “FM-D info, Neues und Wissenswertes aus der zentralen Datenverarbeitung“
Mai 2011)*

Publication amendment form

Begleitblatt für Publikationen des HZB

(Gehören die Autoren zu mehr als einer OE des HZB, ist für jede OE ein Begleitblatt auszufüllen)

OE:		F&E Projekt		Jahr	
Titel					
Autoren mit Adressen					
<u>Nach Erscheinen einzutragen:</u>					
Zeitschrift					
Band					
Seiten von bis					
Federführender Autor					
Antragsteller in der OE					

Vom Antragsteller auszufüllen:

Ich habe keine Zweifel, dass die Regeln guter wiss. Praxis eingehalten wurden.*)
Die folgende Tabelle listet alle - falls ich nur Co-Autor bin, die in meiner Verantwortung entstanden - für die Veröffentlichung relevanten Primärdaten, die Datenträger und die Art der Archivierung:

Daten	Träger	Aufbewahrung

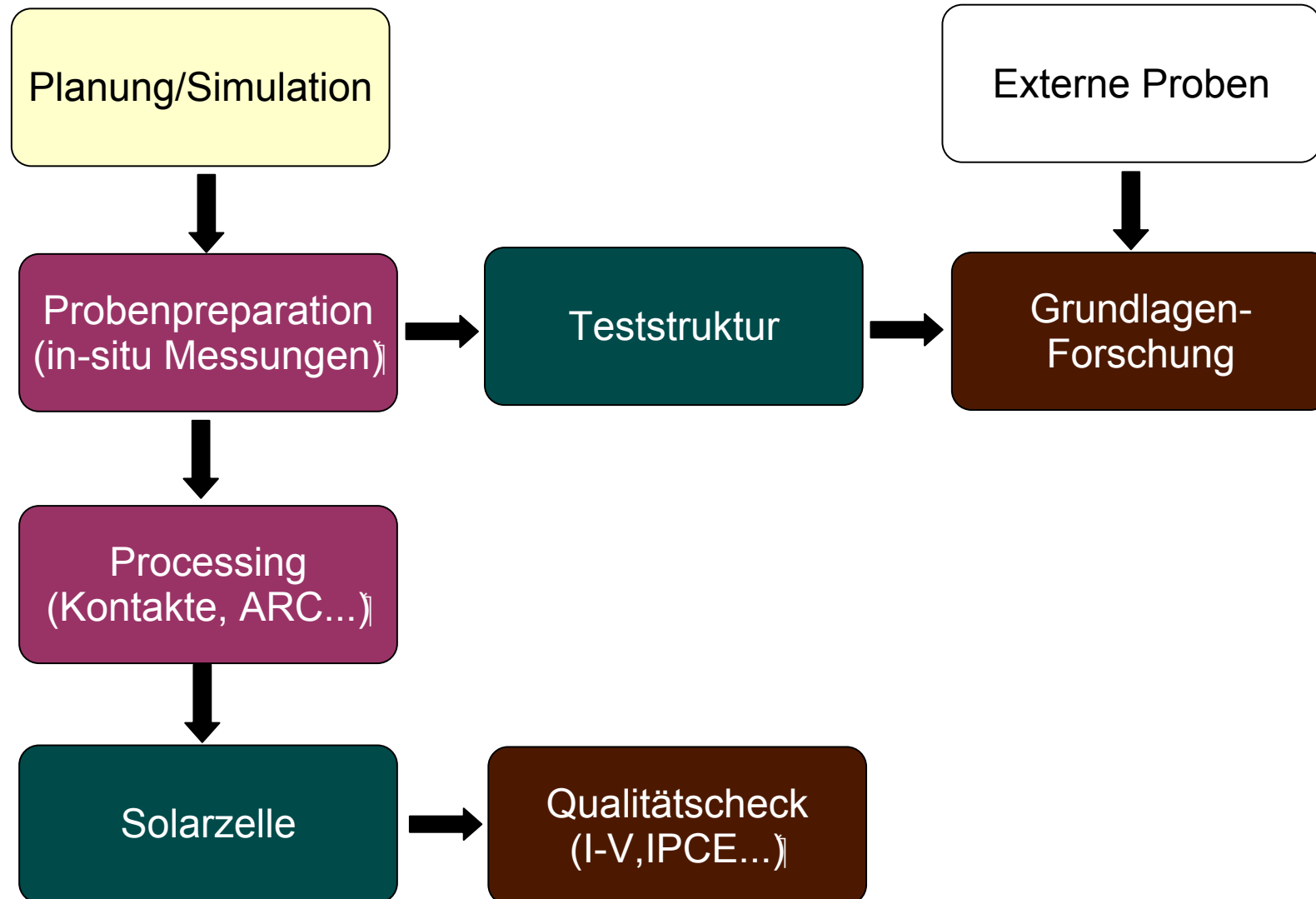
Ich verpflichte mich, alle Träger baldmöglichst dem Archiv zuzuführen, spätestens aber vor dem Ende meiner Tätigkeit als Mitarbeiter des HZB

„The existing rules have been amended, late in 2010, with regulations about orphaned data or rather orphaned-to-be data, i.e. data from scientific organisational units which are being discontinued, in case there is no other scientific organisational unit which can take care of such data. In such case data should be documented and taken care of

- by the library for data in paper form**
- by IT services for digital data“**

(HZB Intranet Bibliothek 23. Mai 2011)

- **Experiments are manifold and complex, i.e. multi-step preparation of materials and thin films (absorbers, buffer and window layers, characterization of materials and devices by a multitude of methods**
- **Data are generated in many different electronic formats or on paper**
- **Volume of data rather small but large amount of variable meta data**



- **Research is largely based on graduate students, large fluctuation makes archiving difficult but especially necessary**
- **No common format or specific rules, quality of archived data depends on individuals**
- **Extra workload for scientists, who often don't see the necessity**
- **Archiving of raw data without link to a publication is often not regarded as necessary nor feasible**

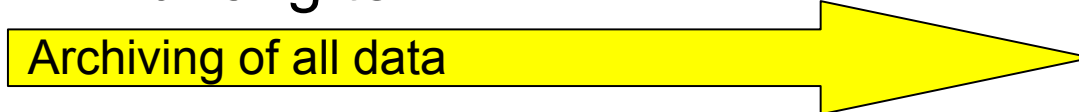
Current

Archiving of publication data



Mid-long term

Archiving of all data



Disadvantage:

Archiving of only publication relevant data is extra 'overhead' because not integrated in a normal work flow of acquiring and automatically archiving data

However, it is debatable if complete archiving of all data and meta data is practical and desirable

Electronic lab book: data table

Experimentdaten Filmherstellung																		
RunNo	Autor	Datum	Quelle No	depositiontime s	Substrate No	Position	Substrate2 No	Position2	Substrate3 No	Position3	intention of experiment	Jod source ml/sec	RunZn source ml/min	PurgeSourceSide ml/min	RunJod ml/min	RunH2 ml/min	Jod push ml/min	Jod pres mba
98070901	gast	09.07.1998	1	60	C1	rot.	C2	7,5 cm	C3	10 cm	Abscheidung bei ca 520 °C Substrattemperatur	400	50	250	15	15	20	600
98070901	gast	09.07.1998	1	3	A1		A2	7 cm	A3	12 cm	Temperaturserie Quelltemperatur 500 °C	400	50	2000	25	25	50	1000
98071301	gast	13.07.1998	7	78	A1		A2	7 cm	A3	12 cm	intention...test	400	50	2000	25	25	50	1000
98071401	gast	14.07.1998	1	3	A1		A2	7 cm	A3	12 cm		400	50	2000	25	25	50	1000
98082601	gast	26.08.1998	1	3	A1		A2	7 cm	A3	12 cm		400	50	2000	25	25	50	1000
98090101	gast	01.09.1998	1	3	A1		A2	7 cm	A3	12 cm		400	50	2000	25	25	50	1000
98091001	gast	10.09.1998	1	60	C1		C2	7,5 cm	C3	10 cm		400	50	250	15	15	20	600

[Excel-Tabelle](#)
[Excel-Tabelle \(deutsches Win95\)](#)

Electronic lab book: search mask

Experimentreihe : Welche Experimente ?

alle Experimente
 RunNo =
 Datum zwischen und
 Experimente mit dem Suchbegriff
 Experimente bei denen
depositiontime zwischen und
 und
depositiontime zwischen und liegt

Experimentreihe : Welche Parameter?

alle Parameter
 Parameter der Gruppe
 ausgewählte Parameter

Charakterisierungen vom Typ

unter dem Namen

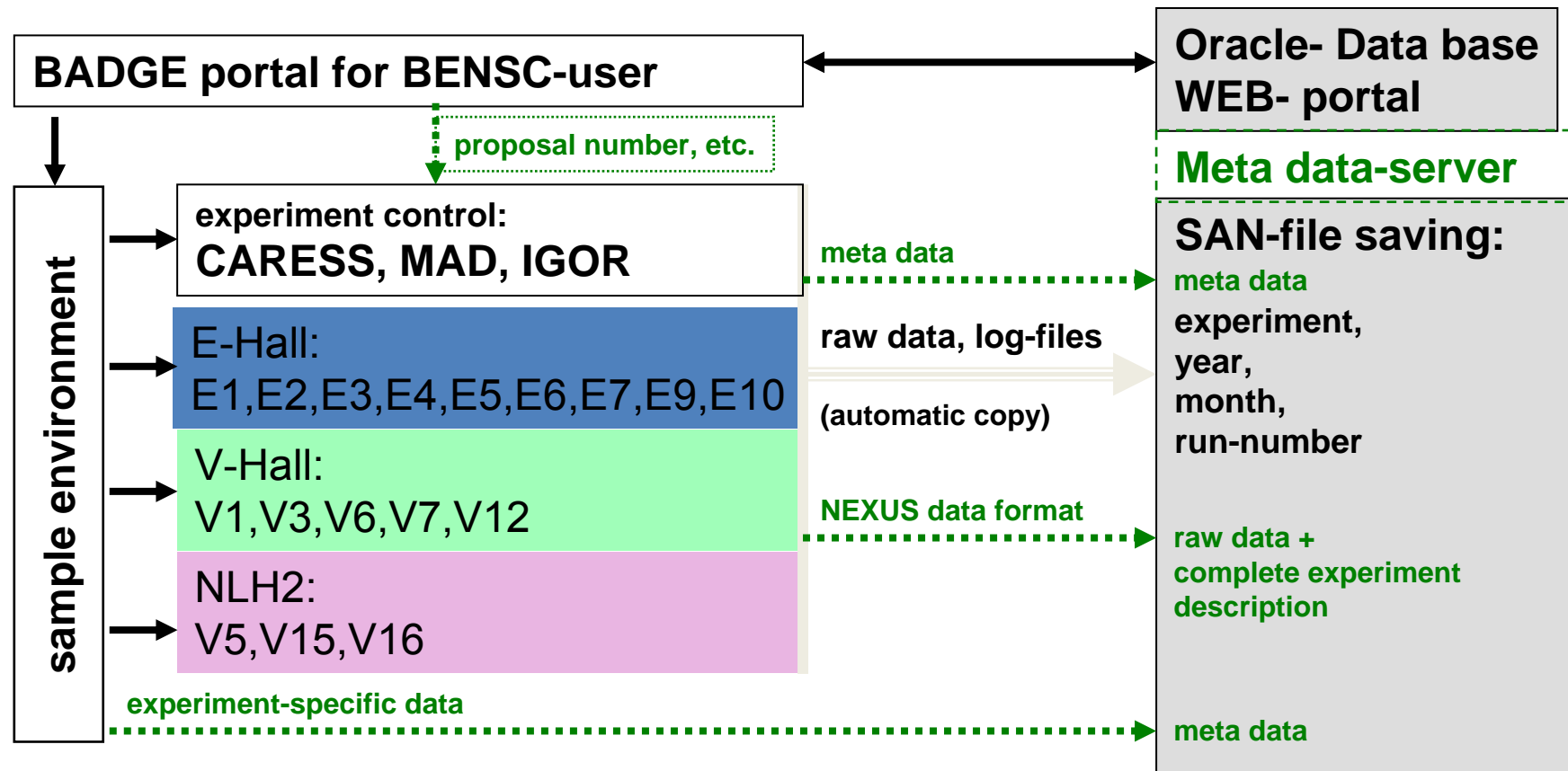
Examples for formats of meta data

Scanned lab book for spectroscopic meta data

Seite	Datum:	Uhrzeit:	Operator(s):	Allgemeines: PES/XES, Probensatz	
8	16.02.06	7:00	J.S.		
Bemerkungen: (Probenhalter, Montage der Proben etc.)				Probenbezeichnung:	Dateiname:
1	XPS	Clase unbeladelt	(S)	1030-7 Pos C	XES160206-007
2	"	"	(S)	1030-7 Pos C	XES160206-008
3		ClS/4mm CAD	(C2)	1041-1 Pos B	XES160206-009
4					
5		Iva/Christian - Herbst			
6					
7	PT # 14	Classe zu MgO, 1.8"			
8	abgeätzt	Pos A: 10s Essigsäure		2474 BB #5	XES160206-010
9	SMU homogenisiert, Spiegelstein geht				
10	in $2,2 \cdot 10^{-6}$ → $1,1 \cdot 10^{-5}$				
11	(RV + 1000) → $2,3 \cdot 10^{-5}$ (RV: 22000)				
12	XES: 11mm	Pos C: 30s Essig			-011
13	Zählrate > 9000, Spalt zu!				
14	XES: 13,8mm	Pos D: 120s Essig			-012
15					
16					
17	PT # 20	KCV-gerade Ref		2474 BB #13a	-013
18	XPS: 11,5	Spalt off / P. Deflekt mit 0,0			-014
19					-015
20	Gold Cu/Zn - Fenster mit XES: 11,5		36°		-016
21	Öhler hilft = 0,4, Probenwinkel 35°				
22	Gold Zn - Fenster bei 36°	XES: 11,5	35°		-017
23	mit Zipping auf 60 (dampflos)		60°		-018
24	Gold bei K=60 auf 1000 V, Det=0,4		60°		-019
25	PT # 11, Pos A	Cu-Zn, XES: 11,5	35°		-020
26	10s Essigsäure				
27	Pos C: 30s Essig		35°		-021

PES	Zeit (s)	Energie (eV)	Gitter	Ordn.	Spalt (mm)	Illumin. (mm)	Slice	Beamline:					Seite
Peak	Bereich	ΔE	E_{pass}	n	Mono (eV) (von Bus)	Harm	Spalt	Gap (mm)	I_{max} (mA)	I_{avg} (mA)	CPS		
X	602	150/2/1	0,3	20	200	1	3mm	16,85	19,4	25,4	30		
2	602	"	"	"	"	"	"	"	14,7	2,4	70		
3	302	1002/1/3	0,3	20	1180	3	3mm	23,02	17,6	4,4	150		
4													
5													
6													
7													
X	150	2/1	0,3	20	200	1	3mm	16,85	19,4	11,4	400		
9											1700		
10													
11													
X	150	"	0,7							230	2000		
13											800		
14	1150												
15													
16													
17											1000		
18	302		0,3								5400		
19	1750		0,7								808		
20	1051	1000/1/3	0,25	20	1200	3	3				800		
21													
22	5:00	360/1/1	0,25	20						19,1	900		
23	6:03	360/1/1	0,25	20							1200		
24	11:09	1000/1/3	0,25	20							1500		
25		1000/1/3	0,25	20									
26													
27	1000	1/3	0,15	20									

current state/ **planned** BENSC-instrumentation



- **Data archiving is necessary and compulsory at HZB**
- **Centralized storage space for paper-based and electronic data is available**
- **So far archiving is done only in the departments on local computers**
- **In most departments only publication-related data archived**
- **Quality of archiving very different and dependent on individuals**
- **Very few common standards, formats and rules established**
- **Improvement by education and better support necessary**

- **Wolfgang Fritsch (HZB library)**
- **Jens Uwe Hoffmann (structure of matter department)**
- **Marcus Kühbacher (structure of matter department)**
- **Christof Rethfeldt (structure of matter department)**
- **Andreas Schöpke (energy department)**
- **Klaus Schwarzburg (energy department)**

And you for your attention!