

# Protokoll

## Workshop

### „Institutionelles Datenmanagement am Institut für Meteorologie“

Datum: 17.12.2012  
Uhrzeit: 16.15 Uhr bis 19.00 Uhr  
Ort: Institut für Meteorologie, Alter Hörsaal  
Teilnehmer TOP 1-6: 27  
Teilnehmer TOP 8: 17  
Diskussionsleiter: Ingo Kirchner  
Protokollführer: Mirjam Hirt

TOP:

- 1) 16.15 Uhr -16.20 Uhr: Einführung von Ingo Kirchner
- 2) 16.20 Uhr -16.30 Uhr: Vortrag von Petra Gebauer: Ergebnisse des Fragebogens
- 3) 16.30 Uhr -16.50 Uhr: Vortrag von Thomas Dümmel: Stadtmessnetz und Operationeller Datenworkflow
- 4) 16.50 Uhr -17.00 Uhr: Vortrag von Ingo Kirchner: Organisation des zentralen Datenpools
- 5) 17.00 Uhr -17.10 Uhr: Vortrag von Daniel Befort: Beispiel eines Datenprojektes am IfM
- 6) 17.15 Uhr -17.25 Uhr: Vortrag von Christopher Kadow: CMIP5 Datenstruktur
- 7) Pause
- 8) 17.40 Uhr -19.00 Uhr: Diskussion

#### ***Zusammenfassung der Diskussion (TOP 8):***

##### **Aktuelles Datenmanagement**

Für jedes Datenprojekt gibt es einen Verantwortlichen/IT-Beauftragten/Administrator, der sich darum kümmern sollte, dass Ordnung in dem Datenprojekt herrscht. Auf Grund fehlender Policies und einer mangelhaften Struktur ist in einigen Datenprojekten (emac, era40,??) dieser Verantwortliche jedoch mit dieser Aufgabe überlastet.

Probleme:

- Häufig werden Datensätze mehrfach neu angelegt und gespeichert.
- Eine Suche nach Datensätzen ist langwierig und führt oft nicht zum Ziel.

##### **Verbesserung der aktuellen Struktur der Datenprojekte**

Die aktuellen Datenprojekte sollten umstrukturiert werden, möglicherweise anlehnend an das Beispiel von Daniel (vgl. Vortrag). Eine Vereinheitlichung für alle Projekte ist jedoch nur schwer möglich.

Es wird vereinbart, dass in einer kleinen Gruppe (Ingo Kirchner, Ines Höschel, Christopher Kadow, Daniel Befort, Markus Kunze und Thomas Bergmann → Policy-Gruppe) ein gemeinsames Konzept zur Struktur der Datenprojekte erarbeitet wird. Zusätzlich sollten alle IT-Verantwortlichen der Datenprojekte, die die Wünsche und Bedürfnisse der Nutzer der Datenprojekte kennen, teilnehmen.

Ziel dieses Treffens ist es, einen Vorschlag für eine neue Struktur der Datenprojekte zu erstellen, bzw. wie die aktuelle Struktur der Datenprojekte verbessert werden kann. Dieser Vorschlag sollte noch im 1. Quartal 2013 bekannt gemacht werden und dem Feedback entsprechend verbessert und umgesetzt werden. Nach der Umsetzung der neuen Struktur sind die IT-Beauftragten dafür verantwortlich, dass diese auch eingehalten wird (Umsetzung selbst kann nicht durch IT-Beauftragte allein erfolgen!).

Anschließend sollten auf diesen Erkenntnissen aufsetzend übergeordnet Policies zum institutionellen Datenmanagement erstellt werden, um in der Zukunft für neue Datenprojekte Richtlinien zu haben.

## **Behandlung von Daten, deren Besitzer bereits das Institut verlassen haben**

Es befinden sich noch viele Daten in dem Datenpool des Instituts, deren Zuständigkeit bei Personen liegt, die nicht mehr am Institut arbeiten. Oft fehlen dabei auch Readme's, weshalb die Daten für eine weitere Nutzung nicht zu verwenden sind.

Da ein Löschen der Daten, die Teil von Veröffentlichungen sind, jedoch gegen die Empfehlungen der DFG zur Sicherung Guter Wissenschaftlicher Praxis ist, wurde folgender Vorschlag diskutiert:

Zuerst sollte eine Liste mit Verzeichnissen, deren Verantwortliche nicht mehr am Institut arbeiten, erstellt werden (Thomas Bergmann), entsprechende Personen sollten kontaktiert werden. Entweder entscheidet der Arbeitsgruppenleiter über die Zukunft der Daten, oder sie werden, sofern sie nicht weiter verwendet werden, auf das Zedat-Back-Up-System ausgelagert.

Um dieses Problem in Zukunft zu vermeiden, sollte ein Übergabe-Protokoll erstellt werden (Policy), wonach jeder, der das Institut verlässt, die Verantwortung für seine Daten an eine andere Person übergeben muss, andernfalls werden diese gelöscht. Dies könnte Bestandteil eines Laufzettels für ausscheidende Mitarbeiter sein.

## **Metadaten und Readme's**

Teil der Policies sollte es sein, dass zu jedem Datensatz ein vollständiges Readme verfügbar ist (muss vom IT-Beauftragten kontrolliert werden). Undokumentierte Datensätze sollten gelöscht werden. Eine Metadatenerfassung innerhalb des Datensatzes (z.B. CMIP5/CMOR) soll jedoch noch nicht implementiert werden, weil die Standardisierung sehr heterogen und in Diskussion ist.

In ferner Zukunft könnten die Ergebnisse des EWIG-Projektes (in Bezug auf Metadaten) jedoch dazu beitragen, eine bessere Metadatenstruktur (im Vergleich zu den Readme's) zu entwickeln und einzubauen.

## **Schulung/Dokumentation**

Damit auch neue Nutzer die Struktur in Datenprojekten besser und von Anfang an nutzen können, ist es notwendig, die existierende Struktur zu dokumentieren und Schulungen anzubieten.

Zwar wurden bereits diverse Dokumentationen erstellt jedoch ist das Wissen, wo welche Dokumentationen zu welchen Daten zu finden sind, unzureichend verbreitet (Lösungsvorschläge: Laufzettel, Default-Startseite,...). Darüberhinaus fehlt das notwendige Feedback, um die Dokumentationen zu verbessern und auf verschiedenen Wegen publik zu machen. Der Aufwand für die Aktualisierung der Dokumente wird gegenwärtig nicht durch die dafür verfügbaren Ressourcen ausgeglichen. Dies erschwert die notwendige Routine, um die Dokumentation auf dem aktuellen Stand zu halten.

Auch sollte es eine regelmäßige Schulung zum Datenmanagement und zur Struktur von Datenprojekten am Institut geben (erstmal Ende 2013), z.B. im Rahmen des Gemeinsamen Seminars für Studenten und Mitarbeiter. Ergänzend zu dieser jähr- oder halbjährlichen Schulung können Videoaufnahmen dieser Schulung und E-Learning-Module für neue Nutzer sehr hilfreich sein.

**Bemerkung:** Auf einem ersten Treffen der Policy-Gruppe am 8.1.2013 wurden Eckpunkte des Papiers erarbeitet. Am 17.1.2013 wurde ein erster Entwurf zur Diskussion freigegeben, der auf einer weiteren Sitzung (voraussichtlich am 8.2.) zur Verabschiedung reifen soll.