

StartClim2004.F

Weiterführung und Ausbau von

MEDEA

**(Meteorological Extreme event Data
information system
for the Eastern Alpine region)**

umweltbundesamt^U



Umweltbundesamt GmbH

Spittelauer Lände 5, A-1090 Wien

<http://www.umweltbundesamt.at>

Martin König, Dipl.-Geogr.

Martin.koenig@umweltbundesamt.at

Abt. für nachhaltige Entwicklung

österreichisches Büro für Klimawandel

Herbert Schentz

Herbert.schentz@umweltbundesamt.at

Abteilung IT-Entwicklung

Katharina Schleidt

Katharina.schleidt@umweltbundesamt.at

Abteilung IT-Entwicklung

IIASA

Tatiana Ermolieva, Dr.

ermol@iiasa.ac.at

Land Use Change Programme

Matthias Jonas, Dr.

jonas@iiasa.ac.at

Forestry Programme

Wien, im Jänner 2005

StartClim2004.F

Teilprojekt von StartClim2004

„Analysen von Hitze und Trockenheit und deren Auswirkungen in Österreich“

Projektleitung StartClim2004:

Institut für Meteorologie,
Department für Wasser-Atmosphäre-Umwelt, Universität für Bodenkultur

Peter Jordan Straße 82, 1190 Wien

URL: <http://www.austroclim.at/startclim/>

Inhaltsverzeichnis

Kurzfassung	4
Abstract	4
F-1 Zielsetzung	5
F-2 Basis Konzept	7
F-2.1 Entwicklung von MEDEA aus MORIS	9
F-2.2 Klassenkonzept	9
<i>F-2.2.1 Basisklassen / Core Ontologie</i>	<i>9</i>
<i>F-2.2.2 Benutzer definierte Klassen</i>	<i>11</i>
<i>F-2.2.3 Polyhierarchie und freie Beziehungen</i>	<i>11</i>
<i>F-2.2.4 Konzeption neuer Klassen</i>	<i>12</i>
F-2.3 Export / Import	12
F-2.4 Selektion	13
F-2.5 Zugriffsberechtigungen	13
F-3 Weiterentwicklung	14
F-3.1 MEDEA-GRID	14
<i>F-3.1.1 Grundkonzept von GRID</i>	<i>14</i>
<i>F-3.1.2 Strategische Bedeutung für die EU</i>	<i>15</i>
<i>F-3.1.3 ALTER-Net / EcoGRID / Austrian GRID</i>	<i>15</i>
F-3.2 Ontologie für Extremereignisse	16
<i>F-3.2.1 Grundkonzept von Ontologien</i>	<i>16</i>
<i>F-3.2.2 Bedeutung der Ontologie für die Vernetzung</i>	<i>16</i>
<i>F-3.2.3 Erstellung einer abgestimmten Ontologie</i>	<i>17</i>
F-4 Literatur und Links	18
Abbildungs- und Tabellenverzeichnis	19

Kurzfassung

StartClim2004.F widmete sich der Weiterführung der Ereignisdatenbank MEDEA (**M**eteorological **E**xtrême event **D**ata information system for the **E**astern **A**lpine region) sowie der Verbesserung ihrer Nutzungsmöglichkeiten für wissenschaftliche Arbeiten. MEDEA soll der Klima- und Klimafolgenforschungsgemeinde in den nächsten Jahren sowohl als Archiv/Datensicherung für mit extremen Wetterereignissen verknüpften Daten in Österreich dienen, als auch längerfristig zentrales Informationssystem für Extremwetterbezogene Daten in Österreich werden, das Auswertungen im on-line Modus ermöglicht.

- ▶ Noch haben die in MEDEA enthaltenen Daten eher exemplarischen Charakter, im Laufe des kommenden Jahres sollen jedoch bereits vorbereitete Vereinbarungen mit wesentlichen Dateneigentümern Forschern den Zugriff auf nennenswerte Datenmengen ermöglichen.

Längerfristig wird der Einsatz von Grid Technologie zur Vernetzung von dezentralen Datenbeständen geprüft. Im Rahmen der Austrian Grid Initiative soll noch 2005 ein Prototyp dieser Vernetzung entstehen, auf dessen Basis MEDEA in Zukunft erweitert werden könnte.

Abstract

The main goal of StartClim 2004.F was the continuation of the meteorological extreme event database MEDEA (**M**eteorological **E**xtrême event **D**ata information system for the **E**astern **A**lpine region). In addition, ways and means were sought after to allow scientists better access to the collected data. MEDEA is designed to provide a long term archive for data relevant to extreme weather events in Austria for scientists working in climatology as well as those working on the consequences of climate change. Current plans include the extension of MEDEA with an on-line interface as well as analysis and modeling tools.

- ▶ Currently, the data integrated into MEDEA has an exemplary character. Based on agreements reached this year with various data providers, the data contained within MEDEA will be extended to allow scientists access to larger quantities of relevant data.

Long range plans for MEDEA include testing the possible use of Grid technology in order to build a network of decentrally stored databases. A simple prototype is in planning for 2005 within the framework of the Austrian Grid Initiative. This prototype could be used as a basis for further development of MEDEA.

F-1 Zielsetzung

Die im Rahmen von StartClim entwickelte Ereignisdatenbank für meteorologische Extremereignisse (MEDEA) verfolgt zwei grundsätzliche Ziele:

1. die vielfältigen Daten derartiger Extremereignisse im ostalpinen Raum zentral zu erfassen und zu sichern;
- und
2. der nationalen sowie internationalen Klimaforschungsgemeinde gezielte Auswertungen aus der Datenbank heraus zu ermöglichen.

Die Bandbreite der StartClim-Projekte sowie der Forschung im Bereich Klimawandel generell zeigt deutlich: extreme Wetterereignisse sind aus wissenschaftlicher Sicht das Spielfeld verschiedenster Fachdisziplinen aus dem natur-, wirtschafts- und sozialwissenschaftlichen Bereich. Während sich MeteorologInnen den Eigenschaften von Wetterextremen (z.B. Ursachen, Prognosen) widmen, sind VertreterInnen insbesondere aus Ökologie, Ökonomie, Raumplanung und Soziologie darauf aufbauend an deren Folgen interessiert (s. Abb. 1). Um diesen Ansprüchen gerecht zu werden wurden im Rahmen von StartClim 2004 folgende Schwerpunkte gesetzt:

- Einbindung von neuen Daten sowie Erweiterung von bestehenden Datenbeständen
- Neue Vereinbarungen bzw. Rechtliche Abklärung bzgl. Nutzung weiterer Datenbeständen
- Erstellung eines Grundkonzepts für eine Internetanbindung um die Inhalte von MEDEA für die StartClim Community leichter zugänglich zu machen
- Recherche bzgl. des Einsatzes von Grid Technologie um eine verteilte Datenhaltung zu ermöglichen

Alle Beteiligten sind auf verschiedenste Arten von Daten zu bereits stattgefundenen Extremereignissen angewiesen, um sowohl für ihre tägliche Arbeit als auch für verschiedene Forschungsprojekte möglichst rasch an Informationen zu kommen. Gerade weil Wetterextreme derart vielgestaltige Eigenschaften und Folgen haben, werden sie in verschiedener Weise von ganz unterschiedlichen Institutionen erfasst, verwaltet und ggf. ausgewertet. Meist sind diese Daten nicht direkt mit einander verknüpft (Beispiel: Wildbachereignisdatenbank des BFW, zentrale meteorologische Datenbanken der ZAMG oder Hochwasserschadensdatenbank des ZENAR). MEDEA soll die technische als auch thematische Verknüpfungen von Daten gewährleisten und damit dem Anspruch genügen, weitergehende Unsicherheits- und Risikoanalysen zu ermöglichen. Die Unsicherheits- und Risikoanalysen sollen auf einem allgemeinen, in sich konsistenten Unsicherheitskonzept aufgebaut werden.

Derzeit ist MEDEA noch als zentrales Dateninformationssystem innerhalb einer Datenbank realisiert. Im Jahr 2005 soll zusätzlich eine Internetanbindung realisiert werden, die es ermöglicht von außerhalb des Umweltbundesamts Daten abzurufen bzw. einzupflegen. Dadurch wird MEDEA einem deutlich breiterem Benutzerkreis geöffnet, wobei der Aufwand der benötigt wird um eine einfache Nutzung zu gewährleisten nicht unterschätzt werden darf! Da es wünschenswert wäre einen Zugriff auf möglichst viele Datenbestände zu realisieren wird eine Vernetzung auf Basis von Grid Technologie derzeit evaluiert. Um einen nahtlosen Zugang zu diesen inhomogenen Datenbeständen zu gewähren wird es nötig sein eine gemeinsame Ontologie zu definieren, mittels dessen die Daten klassifiziert und mit benötigten Metadaten annotiert werden können. Durch den Einsatz von Grid Technologie und Domänenspezifischen Ontologien (Semantic Web) wird durch einen Einstiegspunkt der Zugriff auf alle relevanten Daten gewährt. Ein erster Prototyp diesbezüglich wird im Jahr 2005 im Rahmen der Austrian Grid Initiative erstellt.

Die Ereignisdatenbank MEDEA (**M**eteorological **E**xtrême event **D**ata information system for the **E**astern **A**lpine region) wird vom Umweltbundesamt in Zusammenarbeit mit dem International Institute for Applied Systems Analysis (IIASA) entwickelt und wird allen im Bereich Extremereignisse/Klimawandel tätigen WissenschaftlerInnen und PraktikerInnen als zentrales Dateninformationssystem für Wetterextreme bereitgestellt werden. Das vorliegende Hand-

buch reflektiert den gegenwärtigen Stand dieses Vorhabens, gleichzeitig aber auch, über welche technischen Optionen Benutzer von MEDEA bereits verfügen können.

F-2 Basis Konzept

Bei Start sollten folgende Daten integriert werden:

- Hochwasserdaten des BFW
- Meteorologische Daten der ZAMG
- Agrarökonomische Daten des Forschungszentrums Seibersdorf
- Umfrageergebnisse zu Extremereignissen der BOKU
- Daten zu Tornados des Vereines TORDACH

Bei Sichtung aller Daten, die MEDEA beinhalten soll, wurden vorerst folgende große Gruppierungen der Inhalte gefunden:

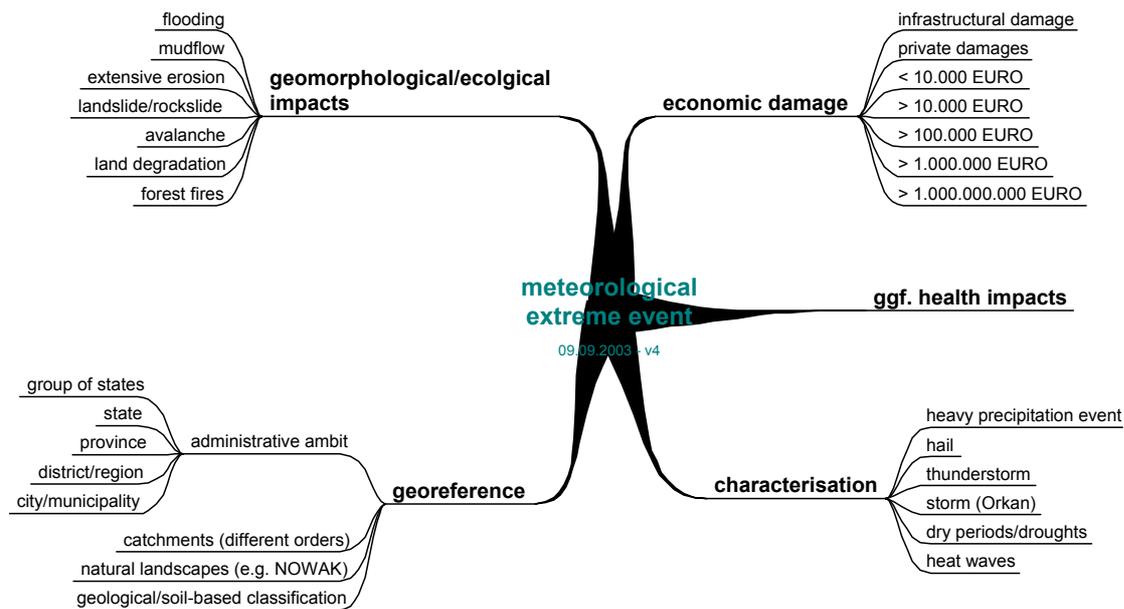


Abb. F- 1 Grundeinheiten

Die dargestellten Entitäten haben alle spezielle Attribute und Parameter, die zudem einige über Zeitreihen hinweg zu beobachten sind. Außerdem gibt es selbstverständlich diverse Beziehungen zwischen den Einträgen, die genauer zu spezifizieren und zu untersuchen sind.

Entität

Eine Entität ist eine in sich abgeschlossene Einheit bzw. Konzept.

Es wird bewusst nicht das Wort „Einheit“ in diesem Zusammenhang verwendet, da dieses Wort allgemein meist mit Maßeinheiten assoziiert wird was zu Verwirrung führen würde.

Großer Bedeutung kommt ferner der Tatsache zu, dass die anfallenden Daten nicht alle von gleicher Sicherheit sind. Abgesehen davon, dass man dies bei Auswertungen berücksichtigen muss, muss zur Qualitätssicherung von Anfang an notiert werden, welcher Sicherheit die Daten angehören.

Sicherheit

Nach dem Unsicherheitskonzept von Moss und Schneider können Daten mehreren Kategorien zugewiesen werden die Aufschluss darüber geben einerseits wie weit es sich um ein allgemein akzeptiertes Konzept vs. einer Theorie handelt bzw. andererseits ob es sich bei den Daten um reine Messwert/Beobachtungen handelt oder ob schon weitere Verarbeitungsschritte durchgeführt worden sind. Daten können nicht nur explizit eines dieser Kategorien zugeordnet werden, sonder befinden sich im Allgemeinen im dazwischen aufgespannten Raum.

Beispiel: Ein Tagesmittelwert ist einerseits eine konkrete Messung (z.B. Temperatur), andererseits durch die Mittelwertbildung wiederum aufgrund eines Modells gebildet

Das „Uncertainty Concept“ von Moss und Schneider ist gut für eine diesbezügliche Kategorisierung der Daten geeignet:

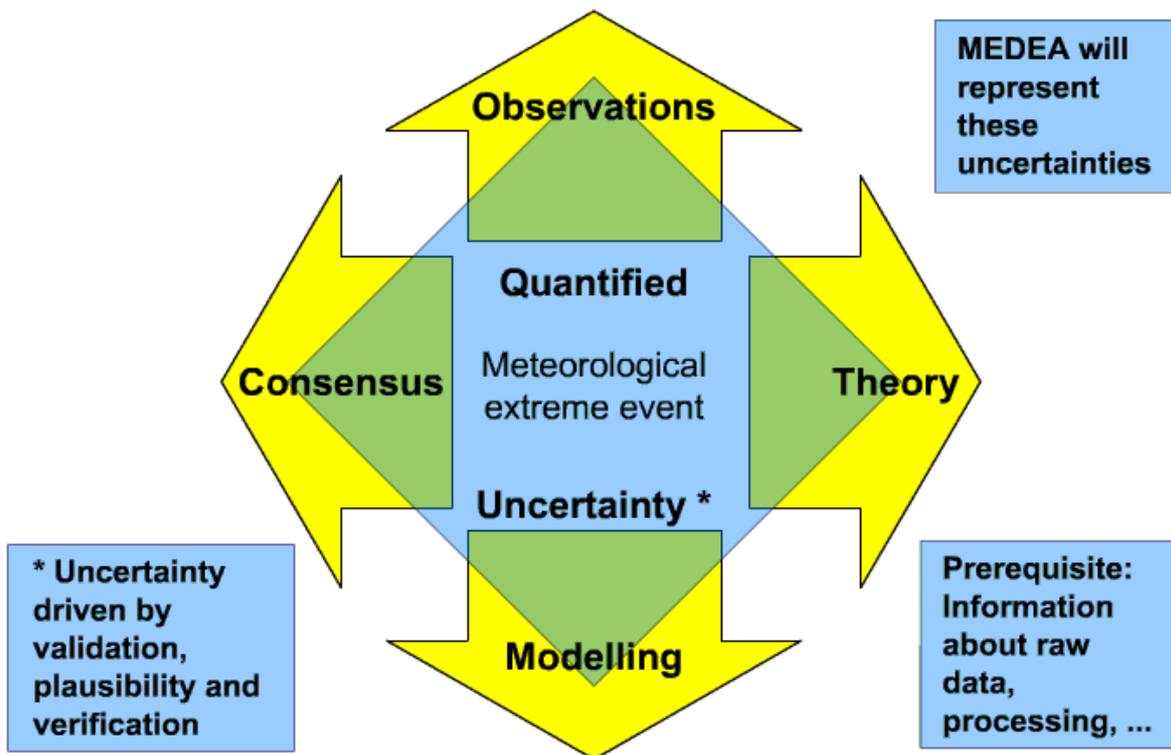


Abb. F- 2 Uncertainty Concept von Moss und Schneider

Es war darüber hinaus von Anfang an klar, dass die in Abb. F- 1 Grundeinheiten dargestellten Entitäten erst den Grundstock darstellen und in Zukunft mit Sicherheit erweitert, feiner spezifiziert werden.

Also gibt es folgende Fixpunkte für das Grundkonzept:

- Es müssen mindestens alle Entitäten aus Abb. F- 1 Grundeinheiten und Relationen zwischen ihnen darstellbar sein.
- Es muss möglich sein, diese Entitäten zur Laufzeit zu erweitern.
- Es müssen Zeitreihen abbildbar und auswertbar sein.
- Die Daten müssen den Unsicherheiten gemäß Schneider und Moss zuordenbar sein.

F-2.1 Entwicklung von MEDEA aus MORIS

Auf der Suche nach einem geeigneten Informationssystem, welches entweder direkt verwendbar, neu parametrisierbar oder zu einem Informationssystem für meteorologische Extremereignisse weiterentwickelbar ist, ist das Team auf MORIS (**M**onitoring and **R**esearch **I**nformation **S**ystem), dem Informationssystem des österreichischen Teils des UN-ECE Projektes Integrated Monitoring, gestoßen, welches nach näherer Analyse alle Anforderungen des Grundkonzeptes erfüllt.

Allerdings gilt es, die MEDEA spezifischen Klassen großteils neu zu definieren.

F-2.2 Klassenkonzept

Das Informationssystem MORIS wurde für jene folgende Anwendungsfälle entwickelt:

- Es ist zwar ein Großteil der Daten vorhanden, man muss aber davon ausgehen, dass sich die Stammdaten im Laufe der Zeit verändern und neue hinzu kommen.
- Die vorhandenen Daten sind heterogen, sowohl in ihren Stammdaten, als auch in ihrer zeitlichen Auflösung, in ihrer Genauigkeit, in der Methode ihrer Gewinnung, usw.
- Alle Metadaten, die primären, wie Auflösung, Nachweisgrenze, die sekundären, wie Erhebungsmethodik und die unstrukturierten wie Skizzen, Photos, Verweise zu Erhebern, usw müssen genau aufzeichnenbar sein.

Das in MEDEA angewandte Konzept von MORIS sieht dafür eine Basis Struktur, heute würde man sagen, eine Core Ontologie (Ontologie = Semantik und Struktur) vor, die vom Endanwender zur Laufzeit gemäß der spezifischen Bedürfnisse erweitert werden kann.

Die Core Ontologie wird durch das Informationssystem an sich repräsentiert und die Erweiterungen werden durch Endbenutzer definierbare Klassen bewerkstelligt. Es können für jede Klasse eigenen Attribute und Beziehungen zu anderen Klassen definiert werden.

F-2.2.1 Basisklassen / Core Ontologie

MEDEA nutzt die MORIS Technologie und übernimmt damit seine Core Ontologie, die durch das Informationssystem verwirklicht wird und sich mehrfach bei Ökologischen Themen bewährt hat.

Diese Basisstruktur ist auch als *.OWL Struktur und als UML Graphik unter <http://www.umweltbundesamt.at/CEDEX> verfügbar.

Klasse

Eine Klasse ist eine Beschreibung der Merkmale (Attribute) einer Dateneinheit.

Beispiel: Die Klasse Person könnte die Merkmale Vorname, Nachname, Geburtsdatum und Telefonnummer haben.

Instanz

Eine Instanz ist ein befülltes Datenobjekt. Eine Instanz beruht auf einer Klasse die angibt welche Informationen diese Instanz beinhaltet.

Beispiel: Die Instanz Mitarbeiter die auf die Klasse Person beruht könnte die Werte „Pepi“, „Huber“, „15.02.69“ und „+43 1 345 67 89“ beinhalten.

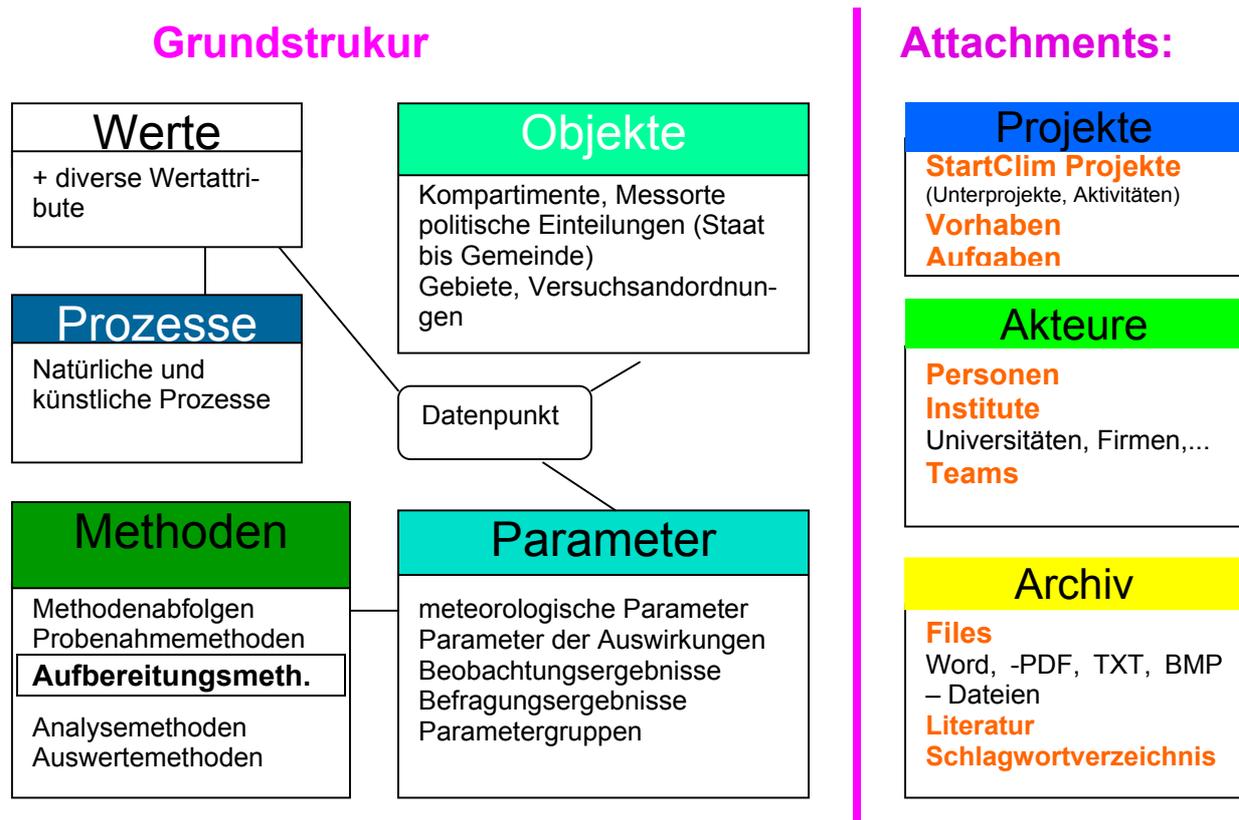


Abb. F-3 Basisklassen von MEDEA

- **Objekt:** Alles, woran gemessen und beobachtet wird, und alle Hilfsstrukturen. Klassischerweise sind Messstellen, beobachtete Gebiete und biotische bzw. abiotische Gegenstände Objekte. Unter den Hilfsstrukturen sind Beobachtungsraster, Einteilungen nach Verwaltungseinheiten, Gebietseinteilungen oder auch z.B. Flusseinzugsgebiete typische Objekte. Messgeräte zählen übrigens auch zu Objekten. Beziehungen sind sowohl zu bestehenden Objekten als auch zu Neudefinierten definiert werden.
- **Parameter:** Alles, was gemessen oder beobachtet wird. Dies können Konzentrationen von Schadstoffen, Fließgeschwindigkeiten von Gewässern, Dicken von Muren, aber auch lediglich nominal einstuftbare Beobachtungsergebnisse, wie Baumzustand bzw. ordinal einstuftbare, wie Lawinenwarnstufe, ... seien.
Für die Überlegung, ob etwas ein Objekt oder Prozessattribut ist, ist, ob der zeitliche Verlauf von Interesse ist. Alles, wovon der zeitliche Verlauf interessant ist, ist über einen Parameter zu definieren.
- **Datenpunkte:** Datenpunkte legen fest, welcher Parameter an welchen Objekten gemessen oder beobachtet wird, wobei wichtig ist, dass nicht alle Parameter, die für eine bestimmte Objektklasse definiert wurden auch an jedem Objekt beobachtet werden müssen. Da man in MEDEA entweder den Parametern ihre Objektklassen oder den Objektklassen ihre Parameter zuordnen kann ist es im Übrigen egal, wie man sich seine Thematik in diesem Punkt strukturiert.
- **Methode:** Die Methodik bestimmt den Parameter. Das heißt, ein mit einer anderen Methode ermittelter Parameter ist ein neuer. Dadurch erhält man zwar mehr Parameter, diese sind jedoch dann eindeutig. Die Polyhierarchie der Parameter erlaubt die Definition von Ähnlichkeiten, was dann auch bei der Selektion wichtig ist. Wesentlich ist, dass alles aus der Beobachtungsmethodik aufgezeichnet wird, vom Untersuchungsdesign bis hin zur Messung oder Klassifizierung. Gerade Daten, die mit anderen verschnitten und verteilt genutzt werden, müssen genauestens beschrieben werden.
Methoden können nicht nur klassifiziert werden, sondern es kann auch noch ihre Abfolge

aufgezeichnet werden.

Ist die Methodik zur Ermittlung der Ergebnisse nicht bekannt, dann muss man sich zuerst sogar gut überlegen, ob man die Daten zur Verfügung stellen kann. Handelt es sich bei der Beobachtung jedoch um einen Parameter, der bereits definiert ist, ist es zu mindestens empfehlenswert einen neuen Parameter mit einer eindeutigen Dummy Methode zu definieren. Es ist nicht gut, Messungen und Beobachtungen bestehenden Parametern zu zuordnen, wenn die Methodik unbekannt ist, da sie zu mindestens eine andere Qualität haben als solche mit bekannter.

- **Prozesse:** Alle natürlichen oder künstlichen Prozesse. Prozesse zeichnen sich dadurch aus, dass sie mehrere Objekte und Parameter, oft recht willkürlich zusammenfassen. Durch den natürlichen Prozess „xxx - Hochwasser“ können recht willkürlich bestimmte Gemeinden betroffen sein. Andererseits werden bei Probenahmen und Kampagnen auch willkürlich bestimmte Parameter an bestimmten Objekten ermittelt. Trotzdem klar ist, dass Prozesse sehr individuell sein können und nicht immer klar definiert sind, ist es gut, sie soweit als möglich zu klassifizieren. Ihnen werden dann Datenpunkte, das heißt an bestimmten Objekten beobachtete oder gemessene Parameter zugeordnet. Dabei kann es sehr einfache Parameter geben wie „ist betroffen“ und genauer gemessen wie „Schadenssumme“, „Lufttemperatur“.
- **Angebundene Information:** Hierbei handelt es sich um „Zusatzinformationen“, die als Vermerke an Beobachtungen und Messungen anhängbar sind. Man kann sie sich sehr gut als Zettel auf der Pin Wand der Objekte, Parameter, Methoden und Prozesse vorstellen.
 - Akteure: Personen, Personengruppen, Institutionen usw. Damit lässt sich sowohl ein Verzeichnis aller wichtiger Personen und Institutionen als auch z.B. Zusammenstellungen von Teams einrichten. Hier ist wohl die wesentlichste Aufgabe, sich die Polyhierarchien innerhalb der Akteure zu überlegen. Die Zuordnung ist später überall möglich. Zusätzlich kann man bei der Zuordnung auch eine Funktion angeben, die man sich vorher überlegen sollte
 - Projekte: Projekte, Vorhaben, Initiativen und ähnliches können hier abgebildet werden. MEDEA ist kein Projektmanagement System, da Ressourcen u.ä.m. nicht verwaltet werden können. Auch für die Zuordnung von Projekten ist es möglich, Funktionen an zu geben.
 - Archiv: Das Archiv kann (noch) nicht mit Klassen gegliedert werden. Aber es ist möglich sowohl Zeiger auf Dokumente oder Dateien ab zu legen, Kurztexte zu erstellen und sie anderen Einträgen und Klassen zu zuordnen. Außerdem gibt es ein Schlagwortverzeichnis und eine Autobeschlagwortung, die für bestimmte Bereiche ein und ausgeschaltet werden kann. So kann z.B. definiert werden, dass Objektklassen automatisch beschlagwortet werden, Objektinstanzen aber nicht.

F-2.2.2 Benutzer definierte Klassen

Innerhalb des oben dargestellten Grundkonzeptes kann der Enduser spezielle Klassen für seinen Bedarf mit eigenen Attributen und Beziehungen zu anderen Klassen definieren.

Damit wird einerseits vermieden, dass das Datenmodell zur Laufzeit geändert werden muss und andererseits wird die Möglichkeit geboten, MEDEA gemäß des Fortschrittes der Wissenschaft und Technik zu adaptieren.

F-2.2.3 Polyhierarchie und freie Beziehungen

Große, aber auch gerade heterogene Datenmengen brauchen vielerlei Möglichkeiten zur Strukturierung. In MEDEA ist die Polyhierarchie ein wesentliches Strukturierungsinstrument.

Hierarchische Zuordnungen sind nicht nur in der IT, sondern auch in wissenschaftlichen Gliederungen im Allgemeinen weit verbreitet. (Ordnung der Verwaltungseinheiten, Artenlisten, Strukturierung chemischer Verbindungen,)

Da Begriffe im Allgemeinen aber nicht nur in einer Hierarchie vorkommen, ist die Polyhierarchie, die zwar IT technisch schwieriger zu verwirklichende, aber der Realität gut angepasste Strukturierung.

Über die Polyhierarchie hinaus gibt es noch die Möglichkeit freier Zuordnungen, welche z.B. für Synonyme, Querverweise, gut verwendet werden können. Sie können im Sinne der Ontologien auch als „Prädikate“ gesehen werden und sind selbstverständlich beliebig erweiterbar

F-2.2.4 Konzeption neuer Klassen

Generell kann als Grundsatz gelten

„So wenig neue Klassen erstellen, wie möglich, so viele, wie nötig.“

Zunächst muss man die gesamte Thematik, die man neu abbilden will, der Core Ontologie zuordnen. Dies ist übrigens ein Grundsatz, der allgemein bei der Erstellung von themenübergreifenden Ontologien gilt, da man vorerst die für alle Thematiken gültige Basisontologie finden muss. In MEDEA ist diese durch das Schema von F-2.2.1 (Basisklassen / Core Ontologie) festgelegt.

Nach Zuordnung des Themas zu der angegebenen Struktur, muss die nächste Frage sein, welche der gewünschten Klasse es schon gibt, und, ob eine vermeintlich neue Klasse nicht ein Synonym für eine bestehende Klasse ist. Man kann die Attribute bestehender Klassen erweitern, nicht aber einschränken. Trotzdem sollte man vorsichtig sein, bestehende, instanziierte Klassen zu verändern.

Solange keine Instanzen zu Klassen eingegeben oder importiert wurden können diese geändert werden.

Über den Klassen gibt es Typen. Bevor man Klassen eingibt oder verändert, muss man sich überlegen, welcher Type diese Klasse zugeordnet werden kann. Typen können Attribute haben, welche an ihre Klassen vererbt werden und sind somit eine Ausweichlösung für die (noch) nicht realisierte Vererbung innerhalb von Klassen.

Es ist empfehlenswert erst bei guter Übung in der Erstellung von Klassen mit der Erstellung oder Veränderung von Typen zu beginnen.

Anfangs und bei Unsicherheiten ist auf alle Fälle zu raten, die bestehende Klassenstruktur, deren Instanzierungen und Beziehungen genauestens zu studieren, bevor man an Änderungen von Klassen und auch von Instanzen geht.

F-2.3 Export / Import

Da man im Allgemeinen Daten als Textdatei oder EXCEL Spreadsheet zur Verfügung hat, ist es wichtig zu wissen, dass diese Daten leicht nach MEDEA importiert werden können.

Die größte Aufgabe besteht meistens darin, die vorhandenen Daten den Konzepten zu zuordnen. Ein Tool, mit dem man die Spalten seiner Tabelle den Klassen und deren Attribute zuordnen kann, steht zur Verfügung.

Es können Klassen, Instanzen, deren Beziehungen, Zeitreihen und Basisdaten importiert und in diversen Formaten exportiert werden.

Für die rasche Übergabe kann man Daten, welche gerade am Bildschirm sichtbar sind, per Knopfdruck an EXCEL und SPSS übergeben.

F-2.4 Selektion

Drei Gründe haben dazu geführt, ein spezielles Selektionstool für MORIS zu bauen:

- Datenselektion sollte für Benutzer auch ohne SQL Kenntnisse möglich sein.
- Benutzer sollten auf die objektorientierte Datenstruktur einfach zugreifen können.
- Einmal getroffene Auswahlen sollten nachvollziehbar und wiederholbar sein, was besonders bei Reportings gemäß Berichtspflichten der EU und anderer Organisationen wichtig ist.

MEDEA nutzt dieses Selektionstool mit Erfolg, weil durch die Speicherbarkeit ein Set von Standardselektionen entsteht.

Jede Auswahl kann aus beliebig vielen Schritten zusammengestellt werden, die über kontextsensitive Operationen miteinander verbunden sind: Hat man z.B. mehrere Parameter ausgewählt, so kann man sich im nächsten Schritt alle Objekte (z.B. Messstellen) anzeigen lassen, an denen der Parameter überhaupt gemessen wurde. Aus diesen kann man die interessantesten aussuchen usw.

Ein bedeutendes Feature ist die Möglichkeit, bestimmte Einträge für die Auswertung gleich zu setzen. MEDEA verlangt, dass alles, was beobachtet und gemessen wird, genauestens dokumentiert wird, was leicht dazu führen kann, dass Dinge (z.B. Parameter), die nicht exakt gleich sind als 2 verschiedene angelegt werden müssen. **Zum Zeitpunkt der Auswertung kann man dann entscheiden, ob die Unterschiede für die betreffende Analyse belanglos sind oder nicht.** Sind die Unterschiede belanglos, werden die 2 leicht differierenden Dinge für die Auswertung gleich gesetzt. Das führt dazu, dass man für die eine Auswertung, wo es nicht notwendig ist weniger genau unterscheidet als für eine andere.

Die ausgewählten Daten kann man sich anzeigen lassen, exportieren, sofort ins EXCEL oder SPSS übertragen oder sich als Zeitreihe alphanumerisch oder graphisch darstellen lassen. Eine GIS Darstellung ist in Arbeit. Bei all diesen Visualisierungen können Größen in umgerechneten Dimensionen präsentiert werden, vorausgesetzt man hat in den Basisdaten solche Umrechnungen definiert.

F-2.5 Zugriffsberechtigungen

Jeder, der mit MEDEA arbeiten will, muss sich identifizieren, muss also als Benutzer angelegt sein.

Für alle Benutzer gibt es 5 mögliche Rollen:

RESTRICTED	darf Inhalte ansehen, aber nicht runterladen
VIEW	darf Inhalte ansehen und runterladen
EDIT	darf Inhalte ansehen und Instanzen eingeben und ändern
SYSEDIT	darf Inhalte ansehen, Instanzen und Klassen eingeben und ändern
MANAGER	darf alles, was nicht durch Integritätsregeln geschützt ist.

Die MEDEA Manager vergeben die Berechtigungen. Es ist möglich, für einen Benutzer mehrere Berechtigungsstufen einzustellen. Klugerweise werden die Manager den zugreifenden Personen auch niedrigere Rollen geben, sodass jeder Benutzer, der hohe Rechte hat, auch „vorsichtiger“ mit niedrigeren Rechten arbeiten kann, wenn er keine besonderen braucht. Dies empfiehlt sich besonders für Personen mit maximaler Berechtigung SYSEDIT und MANAGER.

Bei Bestehen einer Web Version kann man darüber nachdenken, ob man mit der untersten Berechtigung (VIEW) nicht auch anonym zugreifen kann.

F-3 Weiterentwicklung

F-3.1 MEDEA-GRID

Derzeit werden alle verfügbaren, für MEDEA relevanten Daten zentral in einer Datenbank zusammengeführt. Erst dadurch ist es möglich Zusammenhänge zwischen den verschiedenen Datenbereiche aufzuspüren, und in folge ein Verständnis für die Zusammenhänge in der realen Welt zu bekommen. Dieser Zusammenführungsprozess ist aber nicht nur aufwendig und zeitintensiv, er birgt auch folgende Nachteile:

- **Daten werden schnell veraltet:** nur die eingepflegten Werte sind im System vorhanden. Alle sein dem letzten Datenimport aktualisierten oder neu angefallenen Daten sind nicht verfügbar.
- **Großer Datenspeicher ist nötig:** Alle im Netz verfügbaren Daten müssen in der MEDEA Datenbank platz finden. Vorerst handelt es sich nur um einen kleinen Ausschnitt der möglichen Daten, aber wenn alle verfügbaren und relevanten Daten eingepflegt werden sollen wähen eigene Server dazu von Nöten.
- **Datenhoheit:** etliche Institutionen haben bedenken ihre teuer erarbeiteten Daten in eine Externe Datenbank zu geben. Sie haben Angst, dass ihnen die Kontrolle über diese wertvollen Informationen abhanden kommt.

Aufgrund dieser Bedenken wurde die Idee geboren, MEDEA mittels Grid-Technologie umzusetzen. Diese Technologie würde es ermöglichen MEDEA ohne die oben genannten Nachteile umzusetzen.

F-3.1.1 Grundkonzept von GRID

Für die Simulation der Klimaänderung oder Experimente zur Ergründung des Ursprungs des Universums brauchen Wissenschaftler viel Rechenleistung und große Datenspeicher, die von einzelnen Institutionen nicht mehr finanzierbar sind. Aus diesem Grund werden derzeit weltweit Computing Grids aufgebaut - also Gitter von Rechnern, die zusammengeschaltet eine Art Super-Supercomputer ergeben.

Grid-Computing

Grid-Computing ist der Zusammenschluss von Rechenleistungen und Datenbestände über Breitband-Netze. Dafür müssen Rechenaufgaben in Einzelschritte unterteilt werden, um sie von verschiedenen Rechnern abarbeiten lassen zu können. Weiters müssen die verteilt erarbeiteten und abgelegten Daten klar gekennzeichnet werden, so dass klar ersichtlich ist um was für Daten es sich im jeweiligen fall handelt. Sind die einzelnen Aufgaben erledigt, müssen sie wieder zu einem sinnvollen Ganzen zusammengefügt und eventuell visualisiert werden.

Herkömmlich wird Grid Technologie mit aufwendiger Verarbeitung von Massendaten in Zusammenhang gebracht. Ein weiteres Potenzial dieser Technologie ist jedoch die Möglichkeit, dezentral abgelegte inhomogene Daten zusammenzuführen, und für den Endnutzer als eine Große Datenbasis erscheinen zu lassen. Dadurch wird es möglich, durch den Zusammenschluss existenter Datenbestände zusammenhänge zu erarbeiten, und so neues Wissen zu erlangen.

Aufgrund der Tatsache, dass alle Quelldaten weiterhin bei dem Besitzer bzw. Ersteller dieser Daten gelagert werden, kann ohne dem Aufwand eine zentrale Datenbasis zu erstellen auf alle im Netz eingebundenen Daten zugegriffen werden. Weiters kann jeder Datenbesitzer seine Daten weiter pflegen und erweitern, und so seinen Partnern laufend aktuelle Daten zur Verfügung stellen. Weitere Analyse- und Modellierungsschritte können, anhand der aufgefundenen Daten, im Grid durchgeführt werden. Je nach Aufwendigkeit der Verarbeitungsaufgaben können diese auf dem lokalen Rechner ausgeführt werden oder auch auf im Grid-Verbund zur Verfügung gestellte Rechenkapazitäten ausgelagert werden.

F-3.1.2 Strategische Bedeutung für die EU

In Europa wird seit einigen Monaten am EGEE-Projekt der Europäischen Union gearbeitet, das mit vollem Namen "Enabling Grids for E-Science in Europe" heißt. Im Rahmen des Projekts soll ein robustes und sicheres Grid für Wissenschaft und Forschung in allen beteiligten europäischen Ländern aufgebaut werden.

Zweiter Schwerpunkt ist die Verbesserung und Wartung der Grid Middleware, um den Benutzern zuverlässige Dienste zur Verfügung stellen zu können.

Middleware

Middleware macht Rechner grid-fähig. Die Middleware befindet sich zwischen Betriebssystem und Software und sorgt dafür, dass Aufgaben sinnvoll auf unterschiedliche Rechner mit unterschiedlichen Betriebssystemen, unterschiedlicher Software und unterschiedlicher Rechenleistung verteilt werden. Middleware organisiert den Transport komplexer Daten, vermittelt Funktionsaufrufe zwischen den Komponenten oder stellt die Transaktionssicherheit über ansonsten unabhängige Teilsysteme her.

Da Grid-Technologie eindeutig die Zukunft den vernetzten IT darstellt, ist es für die EU unumgänglich den derzeit vorhandenen Vorsprung auf diesem Gebiet zu halten und auszubauen. Deshalb wird in den nächsten Jahren noch viel Forschung auf diesem Gebiet von Nöten sein, wobei die Vernetzung von dezentral gelagerten inhomogenen Daten als weiteres Einsatzgebiet von Grid-Technologie zunehmend Aufmerksamkeit findet.

F-3.1.3 ALTER-Net / EcoGRID / Austrian GRID

Einige Initiativen zur Erstellung von Grid basierende Netze zur Zusammenführung von ökologischen Daten bestehen schon. Hier ein Auszug an relevanten Projekten:

- **ALTER-Net:** dieses „Network of Excellence“, durch das 6. EU-Forschungsrahmenprogramm finanziert, beschäftigt sich mit der Notwendigkeit für die ökologische Forschung, speziell der Biodiversitätsforschung, entsprechende Netzwerke zur Kollaboration zwischen Institutionen herzustellen. Hierbei ist ein wesentlicher Baustein die Erstellung von einem Data-Grid das es ermöglicht auf die bestehenden Datenbestände der beteiligten Institutionen zuzugreifen, und in einem Arbeitsschritt diese Daten zu verarbeiten. Das Umweltbundesamt ist bei der Erstellung dieses Data-Grids federführend.
- **EcoGRID:** dieses Projekt der amerikanischen SEEK Initiative hat sich zum Ziel gesetzt, auf Basis von Grid Technologie einen nahtlosen Zugang zu verteilte Datenbestände zu ermöglichen. Die nötige Metainformation zu den Daten soll im eigens defi-

nierten EML Format vorgenommen werden. Weiters soll mittels Workflow Tools eine integrierte Weiterverarbeitung der Daten möglich sein. Dienste wie zentrale Authentifizierung und Registry Services zum auffinden von benötigten Komponenten runden das Angebot ab.

- **Austrian GRID:** Diese Initiative zur Erforschung von den Möglichkeiten von Grid Technologie wird vom Bundesministerium für Bildung, Wissenschaft und Kultur finanziert. Ziel dieser Initiative ist es die Infrastruktur für ein Österreichisches Versuchsgrid zur Verfügung zu stellen, wie auch neue Anwendungsgebiete zu identifizieren und umzusetzen. Im Rahmen dieser Initiative wurden die Problemstellungen des umweltbundesamts bezüglich vernetzen von Ökologischen Daten sehr interessiert aufgenommen; im heurigen Jahr soll ein erster Prototyp erstellt werden.

F-3.2 Ontologie für Extremereignisse

F-3.2.1 Grundkonzept von Ontologien

Bei übersichtlichen Datenbeständen ist es möglich, anhand von einfachen Listen oder Baumstrukturen relativ leicht die gesuchten Konzepte zu finden, und in folge an die gewünschten Daten zu gelangen. Umso umfangreicher und vielfältiger die Datenbestände werden, umso schwieriger wird es die gewünschten Informationen zu lokalisieren.

Ontologie (Informatik)

aus Wikipedia, der freien Enzyklopädie

Unter einer Ontologie versteht man in der Informatik im Bereich der Wissensrepräsentation ein formal definiertes System von Dingen und/oder Begriffen und Relationen zwischen diesen Dingen. Zusätzlich enthalten Ontologien, zumindest implizit, Regeln.

Ontologien haben mit der Idee des Semantic Web innerhalb der letzten Jahre einen Aufschwung erfahren.

Eine Ontologie lässt sich mit einer Datenbank vergleichen: Struktur (Datenbankschema) und Inhalt (Daten) bilden ein Ganzes.

Probleme ergeben sich z.B. daraus, dass verschiedene Disziplinen einen unterschiedlichen Zugang zu ein und demselben Begriff haben. In anderen Worten, sie würden den gleichen Begriff, je nach ihrem Blickwinkel, anders einordnen bzw. suchen.

Hierbei können Ontologien Abhilfe leisten, da sie es ermöglichen mannigfaltige Relationen zwischen Begriffen zu definieren. Dadurch kann ein komplexes Netz von Konzepten und Begriffen geschaffen werden, wobei es dem Benutzer ermöglicht wird, je nach Betrachtungswinkel auf verschiedene Wege sich der gesuchten Materie anzunähern. Die Relationen und Zusammenhänge die für den Benutzer nicht relevant sind können einfach ausgeblendet werden.

F-3.2.2 Bedeutung der Ontologie für die Vernetzung

So lang sich der Benutzer in einer abgegrenzten homogenen Umgebung bewegt ist es oft möglich mit einfachen Strukturen die vorhandenen Daten hinreichend zu beschreiben. Umso komplexer dieses Umfeld wird, umso unübersichtlicher wird der Datenbestand, und umso komplizierter wird das Navigieren innerhalb dieser Daten.

Wenn verschiedene inhomogene Datenquellen zusammen vernetzt werden, kann es schnell sehr unübersichtlich werden. Es ist oft nicht von vornherein klar, wo welche Daten in welcher Qualität aufliegen, da verschiedene Institutionen oftmals die gleiche Art von Daten mittels jeweils verschiedener Methodiken erheben, was oftmals zu gravierenden Unterschieden in den Ergebnissen führt. Solange das Wissen über all diese Details nur in den Köpfen der Da-

tenhalter existiert, kann es deshalb bei Vernetzungen zu schweren Fehlinterpretationen kommen. Umfassende Ontologien können hierbei Abhilfe schaffen, wenn sie es ermöglichen nicht nur die Art der Daten (z.B. Niederschlagsmenge) zu bestimmen sondern auch z.B. die verwendete Methodik oder die erfassende Person in gemeinsamer Semantik und Struktur anzugeben. Mit diesem Zusatzwissen ist es dann viel leichter sich ein Urteil über die vorhandenen Daten zu schaffen. So wird er erst möglich, sich in so einem breiten Datenbestand zu orientieren. Eine Applikation (z.B. ein Auswertprogramm) kann überhaupt erst durch die maschinenlesbare umfassende inhaltliche Beschreibung Daten richtig umsetzen.

F-3.2.3 Erstellung einer abgestimmten Ontologie

Die in diesem Dokument dargestellten Klassen und Instanzen wurden definiert, um die Themata meteorologischer Extremereignisse, wie sie im Projekt StartClim vorkamen, zusammenzufassen. Es ist der erste Ansatz einer gemeinsamen Ontologie für meteorologische Extremereignisse.

Eine solche gemeinsame Ontologie über unterschiedliche Themata und unterschiedliche Institutionen hinweg ist der Kern für die Zusammenführung von Daten, unabhängig davon, ob diese Zusammenführung nun innerhalb einer Datenbank oder eines Netzes geschieht.

Es ist selbstverständlich, dass eine solche gemeinsame Ontologie diskutiert und laufend erweitert werden muss, um ein Commitment aller Partner, die Daten gemeinsam nutzen wollen zu erreichen. Diese Erweiterung geschieht sowohl wenn neue Arten (Konzepte) von Daten in das System eingebracht werden wie auch wenn der schon erfasste Datenbestand neuen Nutzerkreisen zugänglich gemacht werden soll. Bei dieser Erweiterung muss sorgfältig vorgegangen werden. Es können leicht parallele Konzepte entstehen wenn nicht genauestens geprüft wird, ob ein äquivalentes Konzept schon besteht bevor ein neues eingebracht wird. Weiters ist es von großer Wichtigkeit dass bei der Erstellung der Basisontologie wohlüberlegt vorgegangen wird um möglichst klare Strukturen zu definieren. Die in MEDEA umgesetzte Basisontologie ist weitestgehend von der MORIS Integrated Monitoring Ontologie abgeleitet, die sich schon über viele Jahre erfolgreich bewährt hat.

Genauso selbstverständlich ist es, dass die Definitionen (Konzepte) in andere Sprachen übersetzt werden müssen, um die Zusammenarbeit möglichst vieler (europäischer) Nationen zu ermöglichen. Diese Mehrsprachigkeit lässt sich auch vorzüglich in Ontologien umsetzen.

Das Umweltbundesamt wird bereits Anfang März ein EU-Projekt in den IST-Part des 6. Forschungsrahmenprogramms einreichen, dass die Erarbeitung einer umfassenden Ontologie für Meteorologische Daten zum Inhalt hat. Sollte die Einreichung erfolgreich sein, ergeben sich hier viele Synergien.

F-4Literatur und Links

ALTER-Net

<http://www.alter-net.info/>

Eco GRID

<http://seek.ecoinformatics.org/Wiki.jsp?page=EcoGrid>

Austrian Grid

<http://www.gup.uni-linz.ac.at/austriangrid/>

Abbildungs- und Tabellenverzeichnis

Abbildungen

Abb. F- 1 Grundeinheiten

Abb. F- 2 Uncertainty Concept von Moss und Schneider

Abb. F- 2 Uncertainty Concept von Moss und Schneider
Basisklassen von MEDEA