



TN4 – Datenmodellierung und Wissensstrukturierung

Reiner Borchert

Version 1

Stand: 10.4.2003

ZUSAMMENFASSUNG

Diese Technical Note umreißt die Aufgaben der Wissensbasis im FLUMAGIS-Projekt, definiert ihre Komponenten und beschreibt die für ihre Erstellung verwendeten Werkzeuge.

Inhaltsverzeichnis

1	Einleitung	3
2	Aufgaben der Wissensbasis	3
2.1	Verbindung zu den Geodatenbanken	3
2.2	Plausibilitätskontrolle	3
2.3	Bewertung und Analyse des Ist-Zustands	4
2.3.1	Leitbilder	4
2.3.2	Defizitfeststellung	4
2.3.3	Kausalanalyse	4
2.3.4	Potenzialanalyse	5
2.4	Maßnahmen vorschlagen und umsetzen	5
2.5	Simulation	5
2.6	Prognosen	6
2.7	Wirkungsanalyse und Erfolgskontrolle	6
3	Komponenten der Wissensbasis	6
3.1	Ontologien	6
3.2	Datenbank-Anbindung	6
3.3	Java-Bibliotheken	7
3.4	Infrastruktur für Simulationen	7
4	Methoden und Werkzeuge	7
4.1	Ontologien	7
4.2	Java-Programmierung	8
4.3	Integration weiterer Komponenten	8

1 EINLEITUNG

Ein interdisziplinäres Projekt wie FLUMAGIS versammelt das Fachwissen verschiedener Fachdisziplinen unter einem Dach. Daraus ergibt sich sowohl die Notwendigkeit wie auch die Chance, dieses heterogene Wissen auf neuartige Weise zu verknüpfen, so dass Zusammenhänge deutlich werden, die bei fachbezogener Sichtweise „unter den Tisch“ fallen würden. Es ist eine große Herausforderung, das verfügbare Wissen so zu strukturieren und mit den vorhandenen Daten zu verbinden, dass die Synergie tatsächlich nutzbar wird. In dieser TN wird skizziert, wie dies im Einzelnen geschehen kann.

2 AUFGABEN DER WISSENSBASIS

Die Wissensbasis verknüpft einerseits das Fach- und Methodenwissen mit den Geodaten, stellt zum anderen aber auch eine Schnittstelle zum Anwender dar, weil sie alle Editiervorgänge an den Geoobjekten auf Machbarkeit und Plausibilität prüfen muss. Das bedeutet, dass der Zugriff des Nutzers auf Geoobjekte und ihre Attribute durch Abfragen und Editiervorgänge ausschließlich über die Wissensbasis läuft.

2.1 Verbindung zu den Geodatenbanken

Die Wissensbasis ihrerseits benötigt einen möglichst direkten und schnellen Zugriff auf die Geodatenbanken. Sie macht die Geodaten für den Nutzer verfügbar, indem sie Datensätze und Felder als Objekte und Attribute darstellt.

Editierattribute werden von der Geodatenbank so verwaltet, dass sie dem Szenario, in dem sie verändert wurden, eindeutig zuzuordnen sind und nur in diesem Gültigkeit haben. Gleiches gilt für Attributänderungen im Rahmen von Simulationsabläufen.

2.2 Plausibilitätskontrolle

Nach dem Abschluss eines Editiervorgangs prüft die Wissensbasis, ob die vorgenommene Änderung überhaupt möglich ist. Wenn nicht, wird die Attributänderung zurückgewiesen bzw. rückgängig gemacht.

Optional kann eine Bewertung der Attributänderung erfolgen im Bezug auf die zu erwartenden „Nebenwirkungen“. Hier kann ggf. eine „Warnung ausgegeben werden, wenn mögliche negative Auswirkungen festgestellt werden.

Eine echte Erfolgskontrolle ist allerdings zu komplex und zeitintensiv, um sie nach jedem Editiervorgang durchführen zu können. Daher wird diese Kontrolle im Rahmen der Wirkungsanalyse angeboten.

2.3 Bewertung und Analyse des Ist-Zustands

Die Bewertung des Ist-Zustands ist eine Klassifizierung der Objektattribute anhand von Werteskalen. Die Analyse kann zum einen diese Bewertung durchführen (sofern sie nicht bereits als Datensatz vorliegt), zum anderen aber feststellen, ob die Bewertung auf ein Defizit und damit verbundenen Handlungsbedarf hindeutet.

2.3.1 Leitbilder

Als Maßstab für die Bewertung des Istzustands werden parametrisierte Leitbilder definiert, die den optimalen bzw. unter den gegebenen Umständen bestmöglichen Gewässerzustand beschreiben. Als „Leitbild“ wird der Kürze halber auch der „Gute ökologische Zustand“ verstanden, der sich vom eigentlichen Leitbild lediglich durch eine etwas abgeschwächte Defizitfeststellung unterscheidet.

Ein Leitbild hat jeweils eine regionale Gültigkeit; daraus ergibt sich die Möglichkeit, dass die Wissensbasis selbst auf Grund des zu analysierenden Bereichs das jeweils gültige Leitbild heranzieht.

Im Leitbild sind die für die Analyse relevanten Parameter definiert. Hier sind für jeden Parameter Wertebereiche oder Grenzwerte festgelegt, so dass die Wissensbasis ggf. Abweichungen vom Leitbild feststellen kann.

2.3.2 Defizitfeststellung

Mögliche Abweichungen vom Leitbild werden als *Defizittypen* definiert. Ein Defizittyp enthält die Informationen über den relevanten Parameter und dessen Wertebereiche und Grenzwerte sowie Hinweise auf *mögliche Ursachen* für ein Defizit. Wird im analysierten Bereich eine Abweichung vom Leitbild festgestellt (*Defizitfeststellung*), generiert die Wissensbasis ein *Defizitobjekt*, welches Informationen zu den betroffenen Geobjekten, zum Defizittyp und zum Grad der Abweichung bündelt.

Die möglichen Ursachen für Defizite werden ebenfalls als *Ursachentypen* klassifiziert und den Defizittypen zugeordnet. An die Ursachentypen werden wiederum *Maßnahmentypen* gekoppelt, die Maßnahmen beschreiben, welche geeignet wären, die Defizitursache zu beseitigen.

Durch die hier skizzierte Verkettung ist es möglich, von einem festgestellten Defizit auf die möglichen Ursachen und die Maßnahmen zu schließen, die die Wissensbasis dem Nutzer vorschlagen kann.

2.3.3 Kausalanalyse

Die Defizitfeststellung liefert eine Liste von Defiziten, die im analysierten Bereich gefunden wurden. Jedes Defizit liefert über die Verknüpfung mit einem Defizittyp Hinweise auf mögliche Ursachen für das Defizit.

Die Kausalanalyse versucht, für jedes der Defizite tatsächliche Ursachen zu finden, indem sie in Frage kommende Geobjekte sucht, die dem Ursachentyp entsprechen (Beispiel: Einleiter als

Ursache für organische Belastung im Gewässer). So können auch die vorgeschlagenen Maßnahmen auf die konkreten Gegebenheiten eingegrenzt werden.

2.3.4 Potenzialanalyse

Parallel zu den Defizittypen werden für ein Leitbild *Potenzialtypen* definiert, die Möglichkeiten beschreiben, welche einigen Geoobjekttypen innewohnen (Beispiel: Altarm, kann unter bestimmten Bedingungen an das Fließgewässer angebunden werden). Auch hier sind Maßnahmentypen an den Potenzialtyp gekoppelt, die Hinweise auf die Realisierung des Potenzials geben. Die *Potenzialanalyse* kann dann Geoobjekte feststellen, die ein Potenzial besitzen, und Maßnahmen zur Realisierung der Potenziale vorschlagen.

Ein weiterer Aspekt der Potenzialanalyse ist die *Gefahrenerkennung*. Eine Gefahr stellt ein „negatives Potenzial“ dar; sie geht mit einem möglicherweise in Zukunft auftretenden Defizit einher. Die mit einem negativen Potenzial verbundenen Maßnahmen dienen nicht der Potenzial-Verwirklichung, sondern, im Gegenteil, der Defizitvermeidung.

2.4 Maßnahmen vorschlagen und umsetzen

Als Ergebnis einer Defizit- und/oder Potenzialanalyse schlägt die Wissensbasis Maßnahmen vor, die sich aus den vorgefundenen Defiziten und/oder Potenzialen ergeben und die auf die betroffenen bzw. die verursachenden Geoobjekte angewandt werden könnten. Der Nutzer kann anschließend die Maßnahmen-Vorschlagsliste bearbeiten, indem er sowohl aus den Maßnahmen wie auch aus den Objekten, auf die die Maßnahmen angewandt werden sollen, auswählen kann.

Die Umsetzung von Maßnahmen kann grundsätzlich auch ohne vorherige Analyse und der daraus resultierenden Vorschlagsliste erfolgen. Der Nutzer kann dabei wählen zwischen einer reinen GIS-Funktionalität (Editierung von Geoobjekten, Veränderung der Geometrie etc.) und der Umsetzung konkreter Maßnahmen, die in einem Maßnahmenkatalog verzeichnet sind. Die Umsetzung der Maßnahmen erfolgt über kleine Java-Programme, die unmittelbar an die Maßnahmen-Objekte gekoppelt sind.

Sämtliche Änderungen an den Geoobjekten werden dem aktuellen Szenario zugerechnet.

2.5 Simulation

Die Wissensbasis wird darauf ausgelegt, den Faktor „Zeit“ zu berücksichtigen und die künftige Entwicklung des Flussgebiets zu simulieren. Zu diesem Zweck wird eine Infrastruktur integriert, die in der Lage ist, die zeitliche Abfolge von Aktionen, Maßnahmen und Entwicklungen zu organisieren. Wenn möglich, sollen hierfür vorhandene Tools verwendet werden.

Ob letztlich brauchbare Simulationen erzeugt werden können, hängt sehr davon ab, wie exakt aus dem vorhandenen Fachwissen verlässliche Voraussagen beispielsweise über die Entwicklung der Vegetation abgeleitet werden können. Wichtig ist in diesem Zusammenhang der Umgang mit

unsicherem Wissen, Wahrscheinlichkeiten und Vermutungen. Auch für diesen Aspekt können möglicherweise vorhandene Tools eingebaut werden.

Für den Ablauf einer Simulation ist es unerheblich, ob der Ist-Zustand in die Zukunft fortgeschrieben wird oder ob die Simulation nach der Durchführung von Maßnahmen gestartet wird. Ihr liegt immer der aktuelle Datenbestand des aktiven Szenarios zu Grunde. Der Nutzer legt den Startzeitpunkt (in der Regel die Gegenwart) und den zu simulierenden Zeitraum fest.

2.6 Prognosen

Prognosen basieren in der Regel auf Simulationen; sie sind quasi das Ergebnis einer Simulation. Die beziehen sich auf einen definierten Zeitraum, der sich in der Regel von der Gegenwart in die Zukunft erstreckt.

2.7 Wirkungsanalyse und Erfolgskontrolle

Um die Wirksamkeit von verschiedenen Maßnahmen-Varianten vergleichen zu können, muss zu jeder Variante eine Simulation durchgeführt werden. Die *Wirkungsanalyse* der sich daraus ergebenden Prognosen zeigt dann, welche Variante das bestmögliche Ergebnis liefert. Eine der Varianten sollte dabei immer die „Nullvariante“ (= keine Maßnahmen) sein.

Eine einfachere Version der Wirkungsanalyse ist die *Erfolgskontrolle*, die nach Durchführung von Maßnahmen eine erneute Defizitanalyse durchführt und dadurch prüft, ob das festgestellte Defizit erfolgreich beseitigt wurde.

3 KOMPONENTEN DER WISSENSBASIS

3.1 Ontologien

Das Kernstück der Wissensbasis bilden fachspezifische **Domain-Ontologien**, in denen jeweils die Begrifflichkeit (Semantik) einer Wissensdomäne, so weit erforderlich, abgebildet wird.

Auch die Geo-Objekttypen werden in einer **Objekt-Ontologie** definiert und in deren Klassenhierarchie eingegliedert. Innerhalb der Hierarchie werden die Objektattribute (Eigenschaften und Relationen) an abgeleitete Klassen weitervererbt (auch Mehrfachvererbung ist möglich).

3.2 Datenbank-Anbindung

Die verfügbaren Geodaten liegen in Form herkömmlicher Datenbanken vor. Um sie für die objektorientierte Wissensbasis verfügbar zu machen, müssen die Daten in *Objekte* und ihre *Attribute* verwandelt werden. Dabei kann jede Tabelle der Datenbanken einem *Objekttyp* der

Objekt-Ontologie zugeordnet werden. Die Felder der Tabelle entsprechen den Objekt-Attributen, die Zeilen der Tabelle den einzelnen *Instanzen* des Objekttyps.

Die Wissensbasis ihrerseits benötigt einen möglichst direkten und schnellen Zugriff auf die Geodatenbanken. Sie macht die Geodaten für den Nutzer verfügbar, indem sie Datensätze und Felder als Objekte und Attribute darstellt.

Editierbare Objektattribute werden von der Geodatenbank so verwaltet, dass sie dem Szenario, in dem sie verändert wurden, eindeutig zuzuordnen sind und nur in diesem Gültigkeit haben. Gleiches gilt für Attributänderungen im Rahmen von Simulationsabläufen.

3.3 Java-Bibliotheken

Die aktiven Teile der Wissensbasis werden als Java-Klassen und –Methoden entwickelt. Die Programmiersprache Java ist besonders gut für diese Aufgaben geeignet, da sie vollständig objekt-orientiert ist und sich somit strukturell sehr gut mit den Ontologien kombinieren lässt.

Für die Benutzeroberfläche sind einige Dialogboxen und „Wizards“ erforderlich, beispielsweise für die Konfiguration einer Defizitanalyse oder einer Simulation. Alle objektbezogenen Aktivitäten und Prozesse (Kausale Defizit-Analyse, Maßnahmen-Umsetzung) werden dagegen ganz im Sinne der Objekt-orientierten Programmierung mit den entsprechenden Objekten (Klassen oder Instanzen) der Ontologien verknüpft. Eine Simulation kann somit direkt auf die Methoden (das Verhalten) der beteiligten Objekte zugreifen.

3.4 Infrastruktur für Simulationen

Der parallele Ablauf mehrerer Prozesse und Entwicklungen während einer Simulation erfordert eine übergreifende Infrastruktur, die die Synchronisation sowie den Informationsaustausch der Einzelprozesse ermöglicht. Nur so kann gewährleistet sein, dass Wechselwirkungen verschiedener paralleler Prozesse erkannt werden.

4 METHODEN UND WERKZEUGE

Für das Fachwissen der einzelnen Disziplinen wird eine Wissensbasis aufgebaut, deren Zweck darin besteht, die verfügbaren Objektdaten untereinander und mit dem Fachwissen der beteiligten Fachinstitute zu verbinden. Dadurch sollen komplexe Analysen, Plausibilitätskontrollen, Entscheidungshilfen und Simulationen ermöglicht werden.

4.1 Ontologien

Als „Ontologie-Editor“ wird mit „Protégé 2000“ eine frei verfügbare und kostenlose Open Source Software eingesetzt. „Protégé 2000“ ist an der Universität von Stanford (Kalifornien) entwickelt worden, ursprünglich, um Ontologien für medizinisches Fachwissen zu erstellen. Da es sich aber

von Anfang an an allgemein übliche Ontologie-Standards gehalten hat, ist es in keiner Weise auf die Medizin-Domäne beschränkt.

Ein weiterer äußerst positiver Aspekt dieses Programms ist, dass es kostenlos zur Verfügung gestellt wird, und zwar sogar im Java-Quellcode, das heißt, dass man es selbst verändern und erweitern kann. Aber auch ohne Änderung des Original-Quellcodes verfügt das Programm über weit reichende Möglichkeiten, eigene Module einzubinden. Außerdem können andere Programme, die die Protégé-Java-Bibliothek benutzen, die Ontologie-Projekte bearbeiten. Das bedeutet, dass man alle relevanten Funktionen von Protégé von anderen Programmen aus nutzen kann, ohne die Protégé-Oberfläche dafür öffnen zu müssen.

4.2 Java-Programmierung

Für die Entwicklung der Java-Bibliotheken wird der J-Builder 6.0 als Entwicklungsumgebung eingesetzt. Da der Java-Quellcode von „Protégé 2000“ vorliegt, kann der Ontologie-Editor sehr eng in eigene Entwicklungen eingebunden werden.

Planung und Entwurf der Software erfolgt mittels Together 6.0.

4.3 Integration weiterer Komponenten

Die Integration zusätzlicher Komponenten ist für einige Aufgaben der Wissensbasis durchaus sinnvoll. So ist zu prüfen, ob die „High Level Architecture“ (HLA) als Infrastruktur für Simulationen sinnvoll eingesetzt werden kann, oder ob beispielsweise das Produkt „Netica“ für den Umgang mit Wahrscheinlichkeiten und Unsicherheiten brauchbar ist. Für hydrologische Modelle und Simulationen können Programme wie „Nasim“ verwendet werden.

Alle zusätzlich verwendeten Produkte sollten im Idealfall über eine Java-Programmier-Schnittstelle verfügen, über die sie eingebunden und gesteuert werden können. Wenn diese fehlt, kann notfalls eine spezielle Daten-Schnittstelle neu eingerichtet werden.