Data Integration and long-term Planning of the Observing System as a cross-cutting Process in a NHMS

More than five years ago MeteoSwiss brought the new Data Warehouse System into operation. In the meanwhile this system has become the “enterprise-wide” data integration platform. Besides the centralized metadata repository the Data Warehouse System is the anchor for most data procurement and data delivery processes. This includes large parts of meteorological and climatological products and services. A successful data warehouse comprises organizational issues as well as technical solutions. MeteoSwiss decided to assign the operational tasks of the data warehouse as a cross-cutting activity to the cross-divisional unit “meteorological data coordination”. The presentation will give a review of the existing system including some details about the organization. It will give an outline of the links between the planning and operation of the observing system, the data procurement and the data delivery processes. Finally some of the present challenges will be discussed.
Data Integration and long-term planning of the Observing Systems as a cross-cutting process in a NMS

ECSN Data Management Workshop Nov 4th 2009

Ch. Häberli
Deputy Head Climate Division/Head Meteorological Data Coordination

Contributions from: Ch. Appenzeller, S. Schärer, Ch. Frei, R. Stöckli, R. Erdin, B. Dürr, U. Germann
Agenda

- Challenges
- Basic Strategies, Concepts and Blueprints
- An Example: Improving the precipitation picture
- Consequences
- Summary
The Seamless Time Dimension

- How was the weather?
- Will there be dangerous weather?
- How did climate evolve?
- How will the weather be?
- How will the next season be?

Agenda:
Challenges
Basic Concepts
Solutions
Consequences
Summary
The Seamless Time Dimension

**Agenda**

- Basic Concepts
- Solutions
- Consequences
- Summary

**Challenges**

- Winter
- Summer
Integration of various Data Sources

Observations (remote sensing / in situ) → Output from NWP

Weather and Climate forecasts

Agenda
- Basic Concepts
- Solutions
- Consequences
- Summary

Challenges
Basic Strategy

- MeteoSwiss operates the backbone of the monitoring systems in Switzerland and the nationwide data integration platform for measurements in the atmosphere.

- The data procurement and data processing strategy of MeteoSwiss is an integration strategy.

- Aspects of integration
  - various observing technologies and platforms
  - observing systems outside MeteoSwiss
  - data and meta data

→ comprises all aspects of „Enterprise Data Integration“
+14 pollen
+ 160 Phenological
The backbone of the atmospheric monitoring systems

Very long Swiss snow series (1864-)
long-term trends of days with snowfall ≥ 1 cm
Organisation Concept

Support Processes

Direction

Communication and Coordination

Core Processes

Numerical Weather Forecast

Monitoring Systems

3rd Party Products

Data Integration

Production Warnings

Production Forecasts

Dissemination

Support Processes

Research and Development

Marketing

Customers (Public, Economy, Politics, Research,...)

Agenda  Challenges  Basic Concepts  Solutions  Consequences  Summary
The backbone of the atmospheric monitoring systems
Here, we artificially double the number of gauges by creating for each existing automatic raingauge a hypothetic random daughter-station some km to the east and to the north.
Ex 1: Combination Radar – surface in situ precipitation

**station data**
- accurate values at stations
- coarse resolution (15km, 24h)
- uncertainty between the stations

**radar data**
- high spatial and temporal resolution (1km, 5min)
- uncertainty in absolute values, often bias
Ex 1: Combination Radar – surface in situ precipitation
Ex 1: Combination Radar – surface in situ precipitation

- Predicted Field, 18 August 2005, all, norad: \( \text{rmse} = 0.79 \)
- Predicted Field, 18 August 2005, all, rad: \( \text{rmse} = 0.65 \)
- Predicted Field, 18 August 2005, smn, norad: \( \text{rmse} = 0.92 \)
- Predicted Field, 18 August 2005, smn, rad: \( \text{rmse} = 0.72 \)
Ex 1: Combination Radar – surface in situ precipitation
Ex 2: Combination satellite – surface in situ sunshine

- **Motivation:**
  - SW radiation
    - forces water cycle
    - input variable for modelling the effect of climate change
  - relative sunshine duration:
    - key parameter for long-term monitoring radiation and cloudiness

Hand drawn contours.
Station data + intuition

But we have more information....
Ex 2: Combination satellite – surface in situ sunshine

- Cold air pool cloud layer trapped by topography
- Not strictly a level surface
- Penetration in valleys
Ex 2: Combination satellite - surface in situ sunshine

- Non-contemporaneous information merging

- Similar outset like in climate reconstruction tasks
Ex 2: Combination satellite - surface in situ sunshine

- stations:
  - relative sunshine duration from ~70 stations

- satellite:
  - Heliosat Clear Sky Index (Clearness Index CIN, Cano et al. 1986)
  - 5% (overcast) - 120% (clear)
  - Implemented for the Alps from MSG, special consideration of snow. (Dürr & Zelenka 2009)
  - Part of CM-SAF project
  - Monthly, 2004 - 2008; Resolution ~ 2 km
  - No station data used.
  - Not real-time so far
Ex 2: Combination satellite – surface in situ sunshine

Monthly Relative Sunshine Duration (%) 2004–12

4PCs

NoSat

CIN

Agenda | Challenges | Basic Concepts | Solutions | Consequences | Summary
Ex 2: Combination satellite – surface in situ sunshine

- Combination of satellite and surface in situ sunshine data → added value für gridded data products - also in the pre satellite era

- Satellite data
  - ... introduce plausible details
  - ... improve gridding, almost always measurable
  - ... of special value in special cases
  - ... can not be replaced by geo-topographical predictors

- Method explains 75-95% of the spatial variance during winter (oct...mar) and 40-80% in summer. Mean absolute errors smaller in summer than in winter.
Consequence → monitoring systems

• **seamless time dimension:**
  - lasting operation
  - keep the system adaptive (to new technologies, monitoring needs...)
  - multifunctional systems

• **automation and modernization**

• **various platforms (in situ and remote sensing) → syntactic and semantic interoperability**
Automation and Modernization

Challenge: Increase information content at reduced cost:
replacement of observers by cameras (~ - 10 kCHF/Station-year)
Consequence → data integration

- Keep the integration platform adaptive (allow analyses with data from 1864 till 2009)
- Allow editing for about 50% of the data set
- High importance of meta data
- Data Owner ≠ product owner
- World wide data exchange → data stewardship? Data quality?
- Permanent increase of data volume
- Real-time, 24h/365d, max downtime: 15'
- 1 data source for monitoring the past climate, warnings, forecasts
- Continuous new requirements, technologies, regulations (WMO, ICAO) but no classical requirement life cycle → automation and modernization
Consequence → Organisation
Consequence → Organisation

Main tasks of the cross cutting unit

• Long-term (strategic) planning of the data procurement chains → architecture blueprints and standards for projects

• Support for projects: fitting the projects into the architecture; change management for the architecture

• Support in daily operations: operations coordination of the data warehouse system
Summary (1)

- Maintaining a monitoring system for the atmosphere is a permanent process and challenging cross-cutting process with a strong Integration aspect.

- The main challenge is to establish a rigorous evolution process by:
  1. regular analysis of the requirements (comprehensive!)
  2. regular review of observation capabilities
  3. gap identification and budget review
  4. implementation and operation

for all parts of any „data chain“
Summary (2)

• The strengths of MeteoSwiss' observing system are:
  - the mix of observing technologies and platforms (including the access to data from external sources)
  - the balance between advanced, state-of-the-art and conventional technologies
  - a highly „integrated“ data management system (which includes data starting 1755)
Thank you for your attention
Kundennutzen

- Voll konfigurierbare (d.h. flexible) Aufbereitung von meteorologischen und klimatologischen (und ähnlichen) Daten;

- Intern:
  - Eine einzige Datenhaltung für alle meteorologischen und klimatologischen Daten sowie die zugehörigen Kontextdaten
  - Integration von „externen“ Datenquellen (Bsp Temperatur: ca 300 Messpunkte verfügbar)

- Extern (v.a. Partner aus Bundesverwaltung oder Kantone):
  - Datenhaltung und Archivierung durch MeteoSchweiz
  - Integration von verschiedenen Datenquellen
  - Datenzugriff über Extranet-Applikationen oder Internet
Entwicklungszyklus & Metadatenmanagement

Master Metadaten Repository

Entwicklungsumgebung

Abnahmeumgebung

Produktivumgebung

Failover Umgebung

Dispo Konzept Technik Herausforderungen Organisation Kundennutzen und Erfahrungen
## Implementierung

<table>
<thead>
<tr>
<th></th>
<th>Production</th>
<th>Failover Production (hot standby)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HW</strong></td>
<td>Sun Fire 6800</td>
<td>Sun Fire 6800</td>
</tr>
<tr>
<td><strong>OS</strong></td>
<td>Solaris10</td>
<td>Solaris10</td>
</tr>
<tr>
<td><strong>RDBMS</strong></td>
<td>Oracle10g</td>
<td>Oracle10g</td>
</tr>
<tr>
<td><strong>ETL</strong></td>
<td>PowerCenter 7.1.4</td>
<td>PowerCenter 7.1.4</td>
</tr>
<tr>
<td><strong>Datenbestand</strong></td>
<td>1755 bis aktuell</td>
<td>Letzte 60 Tage</td>
</tr>
<tr>
<td><strong>Max downtime</strong></td>
<td>Max 15’ pro Unterbruch auf Production</td>
<td>(mit Failover Production Betrieb während ca 8 Tagen möglich)</td>
</tr>
</tbody>
</table>
Einsatz von PowerCenter

Dispo Herausforderungen Konzept Technik Organisation Kundennutzen und Erfahrungen
### Projektchronologie

<table>
<thead>
<tr>
<th>Jahr</th>
<th>Ereignis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1999</td>
<td>Jahr-2000-Problem: droht der Daten-GAU?</td>
</tr>
<tr>
<td>April 2000</td>
<td>Expertise zur historisch gewachsenen Datenbankstruktur → DWH Architektur</td>
</tr>
<tr>
<td>2000</td>
<td>Anforderungsanalyse/Konzept</td>
</tr>
<tr>
<td>März 2001</td>
<td>Pflichtenheft Pilot WTO Ausschreibung</td>
</tr>
<tr>
<td>Dez 2001</td>
<td>Audit Proof of Concept (Prototyp = Release 1)</td>
</tr>
<tr>
<td>Dez 2002</td>
<td>Beginn produktiver Einsatz</td>
</tr>
<tr>
<td>2007</td>
<td>alle „legacy systeme“ ausser Betrieb</td>
</tr>
<tr>
<td>2009</td>
<td>Ausbau für Gitterpunktsdaten/GIS</td>
</tr>
</tbody>
</table>
**Spiralförmiges Vorgehen**

**Business Discovery**
Analyse der Geschäftsprozesse

**Start Stufe**
Anforderungsanalyse

**Anwendungs-Entwicklung**

**Kundentest, Schulung**

**Integration/Test**

**Ende Stufe**

**Anwendungs-Design**

**Pilotbetrieb, phase**

**Planung nächste Stufe**

**Rahmenbedingungen**

---

**Dispo**  Herausforderungen  Konzept  Technik  Organisation  Kundennutzen und Erfahrungen
Erfahrungen (1)
## Erfahrungen (2)

<table>
<thead>
<tr>
<th><strong>Tu es</strong></th>
<th><strong>Lass es</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Auf die Daten horchen; breit abgestützte Reviews durchführen</td>
<td>Zu sehr auf einzelne Datenanwender horchen</td>
</tr>
<tr>
<td>Spiralförmig vorgehen &amp; rasch Betriebserfahrung sammeln; Infrastruktur mitwachsen lassen</td>
<td>Alles auf einmal machen (wollen), Wasserfallvorgehen</td>
</tr>
<tr>
<td>Die goldenen Regeln des DB Designs befolgen</td>
<td>Rasch auf Technologie festlegen</td>
</tr>
<tr>
<td>Geschäftsabläufe verstehen</td>
<td>Von Beratern Lösungen aufschwatzen lassen</td>
</tr>
<tr>
<td>Ab Beginn 1 attraktive Applikation zur Datenanalyse bereitstellen</td>
<td>Segregierte Architekturen</td>
</tr>
</tbody>
</table>