

FINDEX – Meta Information System for Substances

Matthias Menger / PetraJahnke

Federal Environmental Agency, Bismarckplatz 1, 14193 Berlin, GERMANY

Tel. +49 30 8903 2037

Fax +49 30 8903 2007

Mail_matthias.menger@ubade

Abstract: FINDEX is a substance meta information system that virtually maps the data in one system. FINDEX is used internally in the German assuring quick access to different in-house information systems and making vital information available for interdisciplinary work. FINDEX offers a ‘one-stop-search’ for all available information without compromising the context of the underlying systems or touching authorization issues. This paper describes concept and benefits of the meta information system.

Keywords: meta information system; share in-house information; access information via metadata.

1 Introduction

Regarding available in-house information for interdisciplinary work the user spends too much time in finding the appropriate information system containing the right data for its purposes. Presupposed the user is aware of all available systems and

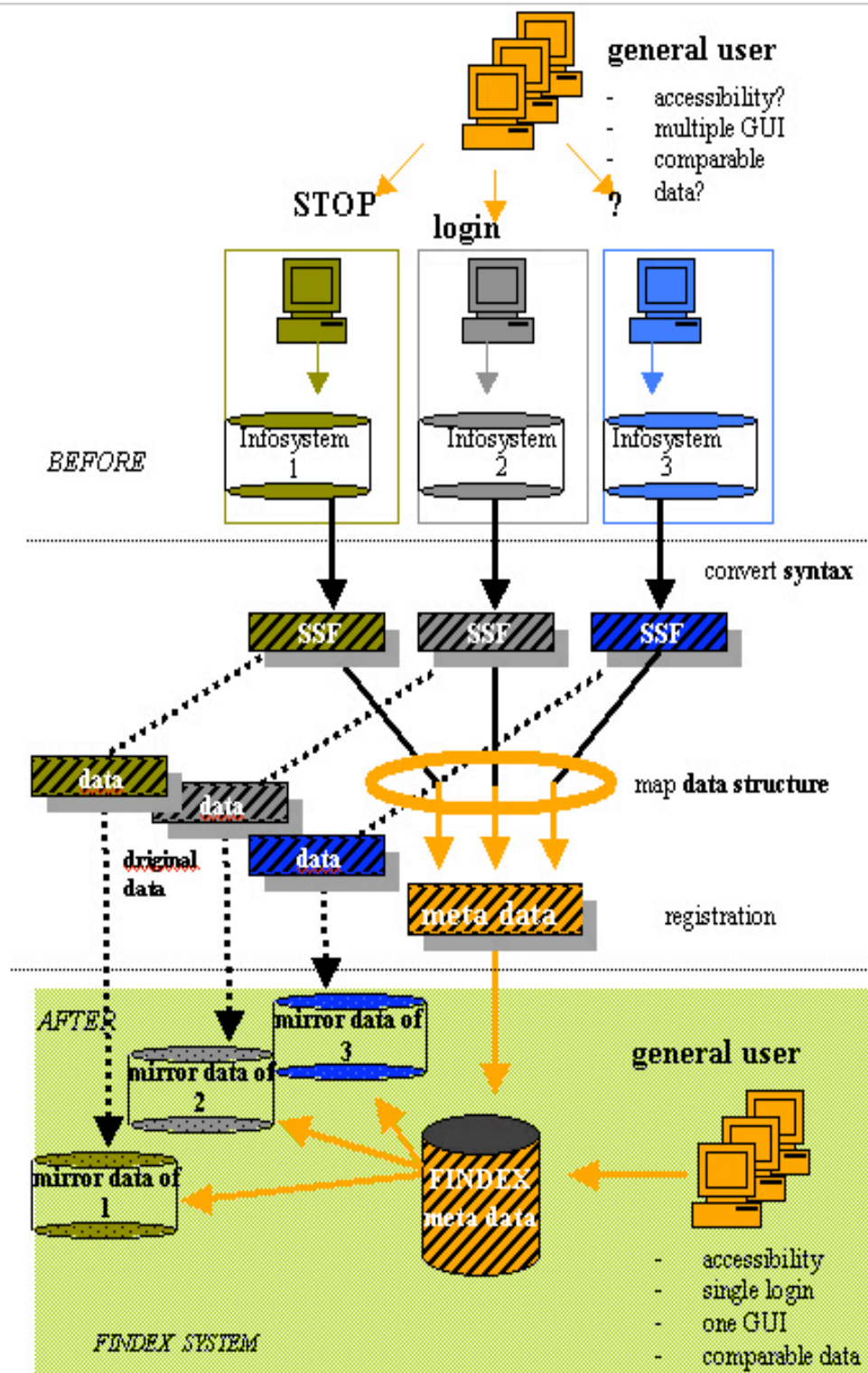
knows where to find and how to access them, which is unlikely in most of the cases. Moreover, he must be authorized to access the specific system due to confidential information. In most cases each system is developed for specific needs of a user group and therefore have not a harmonized data model or even a harmonized graphical user interface (GUI). These facts make it very difficult or even impossible to compare data or information across several systems, especially for substances. FINDEX tries to overcome these barriers and offers a quick overview, based on metadata, of all available information: a single graphical user interface based on a harmonized data model and a direct link to the specific system of interest. Information presented in their ‘original’ context but harmonized for comparison allows the user to decide which information fits best his needs.

2 Metadata System FINDEX

2.1 Harmonization + Metadata

The concept is based on a technical data model which applies following metadata: identity data of a substance (IDENT) and a list of all available characteristics and their occurrences (references) in each of the underlying

information systems. The detailed information are available via these FINDEX-metadata, i.e. the references. Each occurrence provides the direct link to the original data harmonized for presentation in the same application.



This approach assures a first quick overview of all available information

figure 1: Dataflow of FINDEX

The picture shows the procedure and approach before and with the use of FINDEX. A mapping procedure convert the syntax and the data structure of the underlying system into the FINDEX-syntax (so-called SSF) and data structure. For each data model exist a converter, so-called 'cross-reference' (Xref), that transfers the original syntax to the SSF format. Indicating and hiding confidential data is also handled in this converter. The SSF-exports of each information system are mapped afterwards to the harmonized technical data model of FINDEX via Xref (see right part of the picture). This step ensures that substances of the different systems with the same IDENT data (mandatory criteria to identify a substance) are merged via a registration process (see chapter 3) to one information block containing the metadata: IDENT data plus characteristics of the substance and their occurrences in each system.

2.2 Access Underlying Systems

Most underlying systems are restricted to specific user groups due to authorization and confidential issues. Furthermore, some systems are not designed for vast retrieval processes and might be used as 'production system' which should not be opened for 'external' access regarding performance and availability issues (see upper part of the picture: *before*). To overcome this bottleneck of availability FINDEX replicates the data of each system in a performance oriented retrieval format. The already 'syntax-mapped' SSF-exports of each system is transferred into this retrieval

in a single user interface.

format. These so-called 'mirror data' contain all data including the IDENT data in its original context (see left part of the picture). The data are linked to the metadata registered in FINDEX. A frequent update of both the 'mirror data' and FINDEX-metadata ensures up to date information accessible for all in-house users. The original systems are not bothered by 'external' users while their information is available through FINDEX.

3 Flexible Software Approach

FINDEX is based on a very flexible architecture developed in the substance data pool project 'GSBL – Flexible Solution for Substance Information' and therefore offers a potential use for different technical data models. Rule based software modules fit together in a software suite which is controlled by a central handled and adjustable data model.

The registration of substances in FINDEX by their IDENT data is done by a business logic, which identify automatically and unequivocally substances. It distinguish between three types of substances (pure substance, composed substance, substance class) and each follows different defined registration rules. The rules request a name or chemical structure to clearly identify a substance. The logic is able to identify the same substance although its name is mostly written in several different ways, thanks to the step of normalisation: the text string is analysed, separated in sub-strings and normalised to compare it with already existing names. The same logic applies

to compare and register substances via their chemical structure. The logic simultaneously registers a substance and automatically derives information from implicit provided text or the structural data of a substance, i.e. the metadata for FINDEX.

Adopting a different technical data model needs just small adoptions in the defined registration rules of the software, which is currently optimized for substance related information.

4 Benefit of FINDEX

The benefit of FINDEX lies in 'awareness' and the 'accessibility' of all relevant substance information to all in-house users. The original systems keep untouched and need no further authorization, configuration or performance handling.

FINDEX's metadata make stored information aware, show the level of confidence and the availability of information with direct access to the data.

The single graphical user interface and the harmonized metadata give a quick overview and allow comparison of information.

FINDEX provides a flexible system that is able to map every substance related information system and share information for interdisciplinary work. It is a solution to open-up not only legacy and confidential systems for a broader (in-house) use. FINDEX helps to share data, information and knowledge more efficiently. It makes implicit knowledge explicitly available in the agency.

References

1. Feinkonzept
Metainformationssystem zum
Aufzeigen von Stoffinformationen.
Grotz/Roller, Berlin, 1997
2. Stoffdatenrecherche im UBA und
Stoffpflege für den GSBL, Berlin,
Menge, 1998
3. GSBL - Concept + Software Suite of
an Information System for Chemical
Substances, Cottbus, Menger/Jahnke
2003