



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Swiss Confederation

Federal Department of Home Affairs FDHA  
**Swiss Federal Archives SFA**



*corporate information technology*

Swiss Federal Archives SFA  
Dr. Krystyna W. Ohnesorge  
Archivstrasse 24, CH-3003 Bern  
Phone +41 31 32 45827, Fax +41 31 32 27823  
Krystyna.Ohnesorge@bar.admin.ch  
www.bar.admin.ch

itopia ag – corporate information technology  
André Golliez  
Technoparkstrasse 1, CH-8005 Zürich  
Phone +41 44 35 55600, Fax +41 44 35 55601  
andre.golliez@itopia.ch  
www.itopia.ch

# Digital Archiving at the Swiss Federal Archives (SFA)

**Dr. Krystyna W. Ohnesorge, Loris D’Incau, André Golliez**

**The Federal Council decided 2008 to move records management for all federal agencies on a completely electronic basis by end of 2011 and to archive documents created from 2012 onwards only in digital form (GEVER records and process management). In order to be prepared for these changes the Swiss Federal Archives (SFA) developed an innovative solution for archiving digital records and documents which became operative 2009. The core component is the application "Digital Information Repository" (DIR) which accepts digital records and documents submitted by the agencies, stores them on three different locations, provides access to SFA's users and in particular supports the migration approach for preserving digital data across various technology cycles.**

### Initial Situation: the Federal Administration goes Digital

The Swiss Federal Archives (SFA) is the Confederation's service and competence centre for sustainable information management. The SFA preserves records of the Confederation and is responsible for their secure and appropriate safe-keeping, description and dissemination. Records as stated in the Federal Act on Archiving (ArchA, SR 152.1) are "all recorded information, irrespective of the medium, that is received or produced in the fulfilment of the public duties of the Confederation, as well as all finding aids and supplementary data that are required in order to understand and use this information." (ArchA, Art 3, Para 1).

Ever since the mid 1980ies practically all documents and data of the Federal Administration have been produced and processed with the help of electronic information systems. Some have been exchanged and distributed between various offices of the administration via e-mail or have been made available to the public on websites. Paper still plays an important role in legal documents with signatures, in traditional distribution via mail and printed matters as well as in the archiving and preservation field. Furthermore much information is still received on paper at the Federal Administration. The digital world has not replaced the paper (yet) but it takes more and more the leading role thanks to speed, flexibility, intercommunication and allegedly almost unlimited storage capacities. Paper will more and more be regarded as a second tier medium with limited life cycle and limited prospect of its use.

On January 23<sup>rd</sup> 2008, the Federal Council passed a host of measures regarding uniform and standardized handling of electronic data and documents. This action plan is precondition for the implementation of the Swiss E-Government Strategy. In accordance with the Federal Council Decree of January 23<sup>rd</sup> 2008, the Federal Chancellery and the Departments will carry out all their business processes electronically by the end of 2011 (GEVER Confederation Programme). From 2012 on digitally produced records will only be accepted by the SFA if submitted in electronic form.

In order to keep up with this development the SFA elaborated in a first step the strategic and conceptual basic principles on digital archiving within the scope of the E-Government programme and implemented an innovative solution thereupon which became operative mid 2009. This article introduces the conceptual, technical and legal aspects of this solution.

### Tasks and Challenges of Digital Archiving

The secure storing, description and dissemination of the digital records and also the archiving of all other records serve as legal certainty, i.e. the traceability and verifiability of the Confederation's operational activities, the continuous and efficient administration management as well as the prerequisite for a thorough historical and sociological research (ArchA, Art 2, Para 2). In order to conform to this business purpose all archived digital records must comply with the following requirements on a longterm basis, i.e. for an unlimited period of time:

- *Integrity* (through unscathed storing and protection from malicious or unintended mutation and destruction)
- *Comprehensibility* (of the content and the context of the original emergence and use)
- *Originality* (regarding structure and appearance)

- *Authenticity* (regarding authorship and provenience as well as reliability of the obtained evidence)
- *Accessibility* (in legible and reproducible form)

These timely unlimited archiving requirements are opposing the fact that the existence of digital records is always bound to the use of specific information technologies (IT) with a limited life-cycle: specific document and data formats require specific applications to be able to interpret and process these formats, operating systems and hardware to run these programmes. However, the access to content, form and context of digital records has to be ensured far beyond the timely limited life-cycle of the individual IT components. In order to cope with these fundamental challenges in digital archiving, an archive first has to deal with the history, the current state and the future prospects of information technology, and to determine a long-term strategy on how to deal with the technological changes prior to be occupied with its own core competencies. Three alternatives for such a strategy are established at present:

- *Computer Museum*: all hardware and software components where digital objects are based upon are being collected and maintained.
- *Emulation*: any hardware and software components needed for retention and use of the digital objects are being simulated.
- *Migration*: the records will in each case be converted into new data formats or migrated to new media respectively.

The SFA opts for the migration strategy. The enormous technical and organizational effort to maintain a computer museum or emulation procedures is neither economically reasonable nor affordable for the SFA for resources reasons, especially given the fact that the Confederation has a highly heterogeneous IT landscape.<sup>1</sup>

### Conceptual Basis: The OAIS Reference Model

The conceptual basis for digital archiving at the SFA is based on the ISO OAIS Open Archival Information System Reference Model (ISO 14721:2003). The reference model describes an archive as an organization where people and systems collaborate in order to preserve information and make it accessible to a designated community of users. The model describes in detail how information from suppliers ("Producers") is received into the archival system (Ingest), which processing steps have to be carried out under the responsibility of the archival management ("Management") and how the preserved information can be made accessible to the users ("Consumer") (cf. fig.).

The records managing agencies (*Producer*) submit their records to the archive in form of SIP objects (*Submission Information Package*). These objects are integrated into the archive through Ingest functions, i.e. they are converted into AIP objects (*Archival Information Package*) and forwarded to the *Archival Storage* area. Descriptive and administrative information to these archiving objects (*Descriptive Information*, i.e. metadata) are administered and made accessible via *Data Management*. With the Access functions the metadata and the archived objects are made accessible as dissemination objects DIP (*Dissemination Information Package*) to the users (*Consumer*) subject to legal requirement. The area called *Preservation Planning* covers the tasks needed for planning and monitoring the necessary preservation actions. The essential administrative tasks of an archive are being consolidated in the *Administration* area.

<sup>1</sup> Principally there is a (fourth) alternative to store digital records on analog media available requiring only very ordinary technical tools for accessing and processing (i.e. printing everything on paper). Through this procedure, however, the digital character of these archiving objects will be lost and hence not only a significant attribute of their nature but also the possibility of accessing and processing on the basis of the software application. In other words: the archive would remain in the pre-digital period.

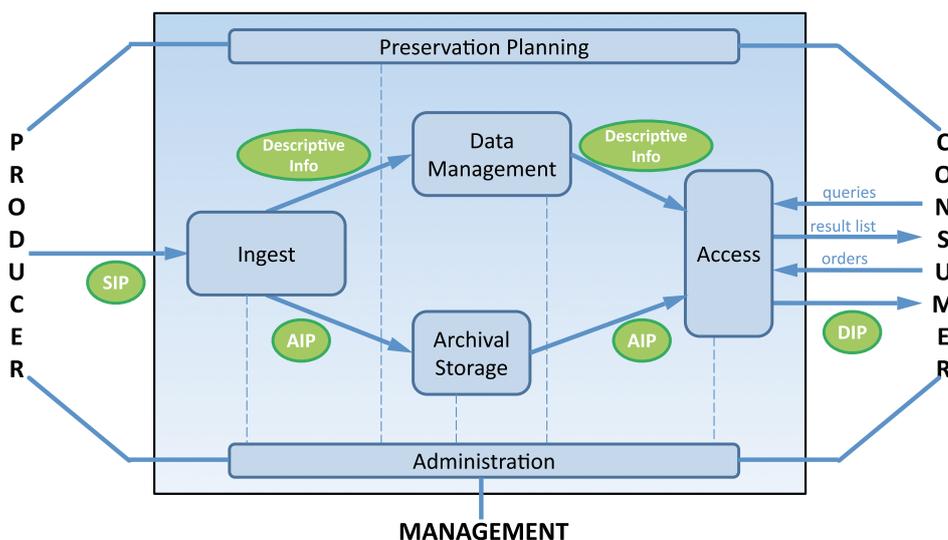
## Legal Basis for Accessing Official Documents of the Confederation

Public access to official documents of the Confederation is regulated on one hand by the Federal Act on Freedom of Information in the Administration (Freedom of Information Act, SR 152.3, since 2006) and by the Archiving Act (ArchA) on the other hand. These two acts build the requirement for a homogeneous practice on the right to access over the entire life-cycle of the Federal Administration's official documents. The access to official documents of the Federal Administration is restricted through the Federal Act on Data Protection (FADP) and the Ordinance of July 24<sup>th</sup> 2007 on the Protection of Information of the Swiss Confederation (IschV).

The Freedom of Information Act has opened the access to originally internal-only records of the Federal Administration to the public in mid 2006 (with restrictions to protect personal information, trade secrets or public

interests). The right to access is guaranteed without claiming particular interests. Exceptions allowing restrictions, postponements of refusals to access are specified according to law.

The Archiving Act (ArchA) follows the principle of free access of the archival data of the Confederation following the expiry of a 30-year safeguard period (ArchA Art 9, Clauses Art 11 and 12). Archival records filed by name and containing particularly sensitive personal data or personality profiles are subject to an extended safeguard period of 50 years (ArchA Art 11, Para 1, OArchA Art 14 Para 1). For access to documents subject to restriction, special authorization is required from the responsible office. For this purpose, a special request through the SFA must be made.



All categories have in common that submissions to the SFA are made in packages containing sets of files with both primary and meta data (and not in single files). As an example a submission from a GEVER system consists of hundreds of dossiers with thousands of documents and thus many files (primary data), the pre-archival classification system (registry plan), the descriptive attributes to every object of information as well as information on the records management agency and the submitting process.

### Six File Formats for Digital Archiving

The limitation of the quantity of file formats is another measure to reduce the complexity in digital archiving. The SFA accepts only a few precisely specified and standardized formats for this purpose:

Specific organizational and technical implementations of an archive compliant to the OAIS reference model has not been specified in the reference model. Especially the OAIS standard cannot be regarded as a specification of the development or procurement of information systems. These conceptual basics have to be elaborated in each archiving organization according to its specific needs.

### Three Categories of Digital Records

Three categories of digital records are being archived at the SFA today:

- **Digital files:** Records from GEVER records and process management systems playing an influential role for the traceability of the Federal business activities. They have to be implemented across the Federal Administration by the end of 2011 (cf. above).
- **Relational databases:** Data generated and maintained within the frame of applications in a specific scope of use.
- **File collections (File repository):** Files generated and maintained in office applications or accrued from photo or video collections.

The concentration on these three record categories is primarily done to reduce complexity. The SFA observes the technological and organizational developments in the Federal Administration and takes account of the emerging requirements of the submitting offices. More record categories will be gradually added in the future, such as geodata (for instance from geographic information systems – GIS).

Application	Format
Text (unstructured)	Plain Text
„Office“ Documents	PDF/A
Tables	CSV
Relational Databases	SIARD <sup>2</sup>
Bit-mapped Images	TIFF
Audio	WAVE

This choice is based on the SFA's practical experience in digital submissions, the knowledge in implemented information systems in the Federal Administration and the estimate that these formats will have a life-cycle of 10 and more years from a present day's perspective on. Such an example is the TIFF Format that has been developed in the late 80ies of the 20<sup>th</sup> century. The choice of formats suitable for archiving and the underlying information and assumptions will be reviewed on a regular basis and new findings and requirements will be adjusted and expanded where appropriate.

<sup>2</sup> SIARD = Software Independent Archiving of Relational Databases, stands for a format for the archival of relational databases that has been developed by the SFA. This format increasingly meets national and international approval and use. The SFA allocate the software solution SIARD Suite for free downloading. SIARD is based on internationally well-recognized standards such as XML, SQL:1999 and UNICODE and presently supports the Oracle, Microsoft SQL Server and Microsoft Access database formats.

### Migration: Preservation through Change

The migration procedures chosen by the SFA aimed at preserving the digital archiving objects throughout several technology cycles include the migration of the archival data from an obsolete file format to a new one. The migration itself is therefore a central and operational function of the archiving solution. It will not be left to a future migration project where it will normally be performed with the replacement of obsolete software products through new applications. Of course quantity and extent of migrations should be reduced as far as possible. A migration should only occur when authenticity, identity, integrity and interoperability of the archival data is guaranteed.

On the analogy of repair or restoration measures to old paper documents, risk considerations have to be applied when migrating digital records. Considerations have to be made whether the possible damage from unwanted side effects during the migration of the archival data is smaller than the damage threatened by the omission of preservation actions. Additional risk reductions can be achieved through storing the digital records both in the old as well as in the new format. Thus every archiving object trails a series of technical versions in different formats over time, whereas normally only the latest records version is available for use.

### Reliable Archive thanks to Recording of Archival Tasks

Design, documentation and replicability of the archival processes are essential for a trustworthy archive. The management of digital records is demanding and needs to be transparent from the moment of submission by the records managing agency. A solution is required that handles every operation on the archival records by a clearly defined workflow and in the recording of archival tasks thereof, be it the examination of a submission, a format conversion or the access to dissemination. Operations on files will not directly be executed by people but through the SFA's own applications exclusively which log every significant operation. Thus the status of the digital archiving objects is documented in a replicable manner, including the technical versions induced by migrations.

### Metadata to the Digital Archive

A submission of digital records consists of primary data in a format suitable for archiving as well as in metadata belonging to them, i.e. descriptive information (administrative, descriptive, structural and technical metadata) that guarantee the research, retrievability and planning as well as the realization of preservation actions (migration) of the records. This includes for example:

- Data on documents (reference number, title, author, etc.) and their order (classification system, file structure, etc.) as well as the origin (records managing agency, systems in use). This information together with the digital records to be archived has to be submitted to the SFA by the records managing agency. Provided that the documents have been administered by a GEVER system the relevant data is usually available and can be exported from the GEVER system via appropriate interfaces.
- Information on characteristics of the files (file name, format, size, check sum, attributes such as the number of pixels for images, etc.). These metadata will be determined and checked during access of the digital records.
- Files that document the management of the archival records by the SFA. These can be simple notes that have been added by the archivists or the description of preservation actions such as the previously mentioned format conversion when performing migrations.

There are many metadata standards. In the prearchival area standards for records management naturally play an important role (International: ISO 15489; Switzerland: see standards of eCH; for the Confederation: the standards of the Federal Strategy Unit for IT FSUIT). There are just as many standards available for metadata in the archive (ISAD(D), ISAAR(CPF), METS, PREMIS, Dublin Core, EAD, EAC, etc.). The various standards often have similar aspects but show differences in the details.

The metadata model for digital archiving at the SFA is based on the ISAD(G) standard (International Standard Archival Description (General)) and depends on the FSUIT's Interoperability Standard I017 in order to integrate the metadata from the GEVER systems of the Federal Administration as simple as possible.

### Administration of Metadata

The term metadata in archiving was coined by information technology. The archivists talk about the description of the records or descriptive information. The SFA has used an Archival Information System (AIS) for years in order to manage the descriptive information. The following table explains the essential differences in the administration of the metadata between AIS and DIR for a better understanding:

	Archival Information System (AIS)	Digital Information Repository (DIR)
<b>Degree of Coverage</b>	Covers all records of the archive, independent of form, state and place.	Covers only those records located in the digital repository.
<b>Depth of Description</b>	Description usually to individual dossiers, seldom to a deeper level.	Description of digital records down to each single file.
<b>Plan of Record Groups at the Archive (tectonics)</b>	Manages the arrangement of the whole archive and rearranges the individual submissions in a factual, organizational and timely arrangement.	Only the inner arrangement of the individual submissions is registered. The submissions will not be interrelated with each other; there is no super ordinate structure.
<b>Loan</b>	Manages the information for lending out the records, including condition and retention periods.	Does not manage data for lending.
<b>Location &amp; Condition of the Records</b>	In case of digital records only a logical reference towards DIR is managed. In case of analogical records both the location and the conservational conditions are managed.	Manages the physical storage location (redundant storage) and the technical aspects to the conservational condition for every file or every digital record.

### Technical Solution Requirements

Based on the conceptual fundamentals mentioned above the following requirements for archiving digital records can be deduced:

- The support of the migration process to preserve the digital records
- The treatment of the archival data as packages with multiple/many files and the corresponding pre-archival metadata
- The administration of the technical metadata
- The support (workflow) and logging of archival tasks

The archiving of digital records is a relatively new field. Only a small offering of IT solutions for the preservation of digital records beyond the period of a technology cycle exists. Further more the requirements may vary depending on the type of institution or may be even contradictory. As an example also a library preserves documents in digital form, however, these can be made immediately accessible to the public as they are public by definition. As opposed to this the access to the documents of an archive is subject to legal regulations such as the Archiving Act (ArchA) or data protection. In particular the technical solution for digital archiving at the SFA has to take the compliance of these legal aspects into account (cf. box).

### The SFA's E-Archive Solution

The SFA uses technologies with a certain maturity and dissemination for the development of a solution for digital archiving. Depending on the assignment the SFA also takes on the role of an innovator or an early adapter of solutions.

The technical solution for digital archiving at the SFA is based on the solution of The National Archives of UK (TNA) in London. With the collaboration of a specialized company from Oxford TNA developed an OAIS-compliant solution for the archiving of digital records that has been implemented in other national archives and libraries in the mean time. The SFA could build upon this solution and has continued the development thereof. The result of this is the application DIR (DIR = Digital Information Repository). With the help of this application the digital records are ingested, secured, preserved and disseminated based on workflows.

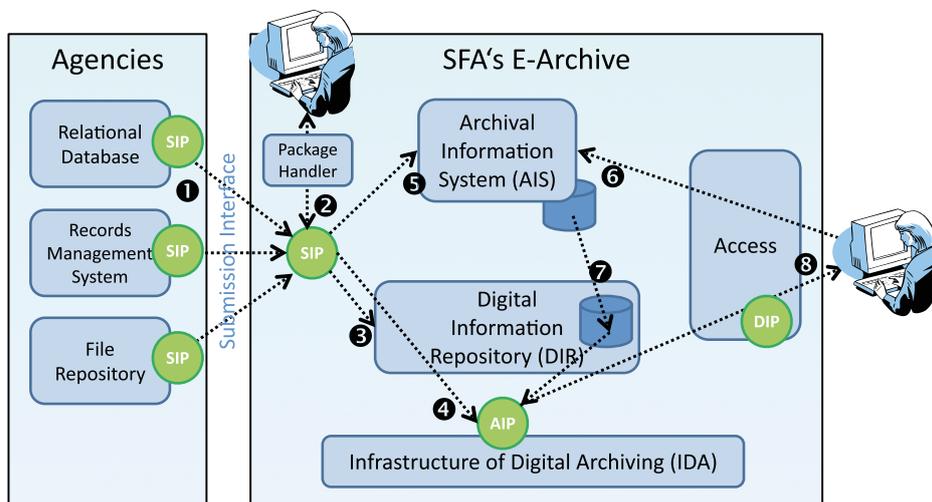
Another important component of the SFA's digital archive is the so-called "Package Handler". The Package Handler is a userfriendly PC-application made for the handling of the OAIS-compliant packages of the digital archive (SIP, AIP and DIP). Employees at the SFA use the Package Handler to browse through primary data and their corresponding metadata in the course of quality checks and to perform modifications on the metadata such as amendments or corrections thereof (cf. fig). The Package Handler is currently being redeveloped so that employees of the records managing agencies are able to independently compile their own submissions according to the requirements by the SFA, if no automatic interface (as for example out of GEVER systems) is available.

For the archiving of relational databases the SFA supplies an additional software solution called "SIARD Suite" that converts relational databases into the SIARD-Format.

The digital records are stored in three different locations within the scope of the "Infrastructure of Digital Archiving" (IDA). As a protection against catastrophes the distance between two locations must be 30 kilometers at least. This infrastructure is operated in collaboration with the Federal Office of Information Technology, Systems and Telecommunication (FOITT).

### Application Architecture

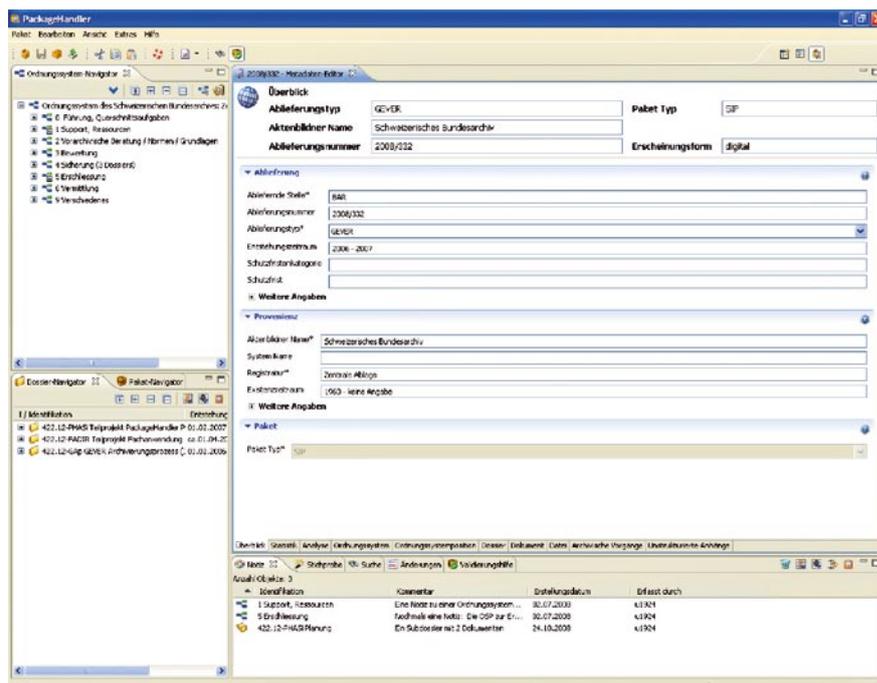
The graphic below shows a simplified application architecture for the archiving of digital records at the SFA.



### The Functionality of the SFA's E-Archive

The functionality of the SFA's E-Archive is illustrated in the following items referring to an exemplary submission of digital records from a GEVER records and process management system and their integration into the digital archive (items 1 to 5) including the dissemination of the archival records (items 6 to 8):

1. It is principally the submitting agency's responsibility to prepare a submission in a form suitable for archiving, also for digital records. The agency



<sup>3</sup> The SFA published the specifications of the submitting objects (SIP) containing all standards needed for the implementation of a SIP when submitting digital records to the SFA (design, structure, content and file formats). These specifications are the standard for software producers of standard solutions within the GEVER records management area.

extracts the closed dossiers that are to be archived including their available metadata from the GEVER records and process management system (disposition) according to their archiving and selection policy, and makes a package as specified by the SFA (i.e. a SIP (Submission Information Package)<sup>3</sup> according to the OAIS Reference Model) (if necessary by using the above-mentioned Package Handler).

2. The SIP will be transferred to the SFA by the submitting agency, whereas different data transfer methods can be used depending on the size and confidentiality of the submissions.
3. The SFA registers the digital submission (SIP) upon arrival within the Digital Information Repository (DIR) and the Ingest workflow will be started (according to the Ingest workflow by the OAIS Reference Model).
  - a. Employees at the SFA check the submissions in view to archiving and legal characteristics (are the submitted records in accordance with the previously agreed upon appraisal? Do the mentioned safeguard periods make sense? Random tests with regard to the completeness of the metadata of the submitted records, etc.)
  - b. DIR checks a series of technical aspects (such as virus scans, actually used file formats, the existence of all referenced files, verification of check sums, etc.).Should the submission not meet the quality requirements, either the whole submission needs to be rejected or the SFA takes over the necessary amendments in consultation with the records submitting agency.
4. As soon as these tasks are completed the definite submission of the package into the archive takes place, i.e. an AIP will be created out of the SIP (according to the OAIS Reference Model, AIP stands for Archival Information Package) and is stored in three copies in the Infrastructure of Digital Archiving (IDA).
5. The metadata necessary for the description in AIS (Archival Information System) will be extracted from the package and will be integrated both into DIR and the Archival Information System (AIS). As a rule all dossiers, but not all individual documents, will be described (cf. box) in AIS. For later identification the described objects will be furnished with distinct reference numbers (UUID = Universally Unique Identifier).
6. A user searches for records in finding aids of the SFA's funds. These finding aids are provided by AIS either via the web or in the reading rooms. If the associated dossiers have been identified in AIS an order can be activated online.
7. The order will be checked in regard to the safeguard period. Should the records not be freely accessible a request for access is needed. The request will be forwarded from the SFA to the records managing agency and the agency will approve or reject it. If the records can be made accessible to the user, the necessary dossiers thereof will be identified in DIR by the distinct reference number out of AIS.
8. DIR compiles the required digital records in a new package (called DIP, according to the OAIS Reference Model a Dissemination Information Package). Depending on the case this package will be made available to the user in the reading room or via the web.

## Conclusions and Outlook

The solution for digital archiving presented here is based on an intense collaboration of several years between archivists and computer scientists. The SFA could build on the most advanced know-how worldwide in this field, thus complement it with own developments in important areas. The focus has always been on the subject-specific requirements of the archive which the technical solution had to comply with. A pragmatical and iterative approach from the concept to the pilot and to the implementation phase as well as the continuous inclusion of submitting agencies stood the test. With this the SFA is well equipped for the highly increasing number of digital records coming from the Federal Administration in the future.

The web-based access to finding aids and archived digital records will open up new possibilities to the users of the Swiss Federal Archives in the near future. Since January 2010 research in the digital indexes can be made online. Records can also be ordered via [www.swiss-archives.ch](http://www.swiss-archives.ch) at the same time. Indexes to 1.2 million archived files from the fonds of the following departments are accessible online: Federal Department of Justice and Police (FDJP), Federal Department of Foreign Affairs (FDFA) (without the Swiss Representation Abroad) and the Federal Department of the Environment, Transport, Energy and Communications (DETEC). During 2010 the indexes of the archived records from the other departments will be amended.

The operationalization of digital archiving as an industrialized service will be in the fore to the Swiss Federal Archives, alongside the IT-assisted access and an expansion of the solution towards additional record categories and file formats during the next years. The quantity of submissions of electronic files from GEVER records and process management systems will increase distinctively in the future and this can only be mastered with a substantial automation of the submission and integration processes. The digital services of the SFA's E-Archive will relieve the Federal Offices from the costly tasks of disposition, converting and preserving their digital records, thus leading to a continuous records lifecycle without the actual breaking point between administration and archive.

### Links

Documentation on Digital Archiving at the SFA:

<http://www.bar.admin.ch/themen/00772/00811/index.html?lang=en>

The SFA's Freeware Solution for the Archiving of Relational Databases:

<http://www.bar.admin.ch/dienstleistungen/00823/00825/index.html?lang=en>

Electronic Administration Management / GEVER Records and Process Management:

<http://www.bar.admin.ch/themen/00697/index.html?lang=en>

Online Research at the Swiss Federal Archives:

<https://www.swiss-archives.ch/suchinfo.aspx>