A SCALABLE INTERACTIVE TV SERVICE SUPPORTING SYNCHRONIZED DELIVERY OVER BROADCAST AND BROADBAND NETWORKS

U. Rauschenbach[1], W. Putz[2], P. Wolf[2], R. Mies[3] and G. Stoll[3],

[1]Siemens AG, Munich, Germany
[2]Fraunhofer IPSI, Darmstadt, Germany
[3]Institut für Rundfunktechnik, Munich, Germany

ABSTRACT

This contribution focuses on two aspects of innovative TV services delivering a new, personalized and interactive rich media experience. The first aspect, service scalability, caters for terminal heterogeneity with respect to screen size, processing power and network connectivity. This way, it allows consuming a rich media TV service using different mobile terminals in addition to the TV set. The second aspect, synchronized content delivery over co-operative DVB and broadband IP networks, provides a flexible way to transmit different components of a service using different network technologies. Depending on content properties, network parameters and current number of subscribers, the best-suited transmission mechanism is chosen automatically for each content item. The system provides end-to-end support to present two media streams synchronously at the receiver side, even if they are transmitted via different networks. As an example a personalized news service is presented which takes advantage of the described features.

INTRODUCTION

In addition to transmitting classic linear TV programs, digital television systems based on DVB provide the foundation to create innovative television services delivering a new, personalized, interactive, rich media experience.

Complementing the TV set, additional mobile devices can be used to consume these services. Recently, the combination of broadcast and multimedia computing devices has generated a lot of interest, see e.g. [1]. For example, a TabletPC with its large screen resolution provides for interactive portable in-home use, and a PDA allows convenient mobile access both in-home and on the move. As these devices have different capabilities (e.g. in terms of screen space and decoding power), scalability of the provided services becomes a crucial issue.

Furthermore, increased bandwidth is becoming available in IP networks, and broadcasters already provide additional Web content to complement their programs. Co-operative usage of DVB and broadband IP networks will not only pave new ways to include additional and personalized content into TV services, but will also lead to cost advantages by using the most appropriate network technology to deliver a particular piece of content. The decision on which network to use has to be done automatically based on characteristics of the content, the networks and the foreseen or even actual usage figures. As some of the additional content (e.g. a signer or additional camera angles) has to be synchronized with the main TV program, an end-to-end solution is required to achieve this when the additional content is not transmitted via the same transmission mechanism as the main TV content.

This contribution proposes solutions for service scalability, smart routing of content over different networks and synchronized presentation. It describes the technical approach taken by the IST SAVANT project[1], which will present a prototype system at the IBC2004 exhibition as part of the EBU village in hall 10, booth 10.411.

SCALABLE SERVICES

A TV program can reach a greater audience if it becomes possible to overcome the time, location and content restrictions imposed by the traditional time-linear TV viewing pattern. A service in the broadcast domain is currently made up of a set of time-related content components, containing visual and audio information. Programs are designed to be *immediately consumed* in a *time-linear* fashion on *TV sets only*.

A News program is a good example where enhancements would be of great value for users as they are interested in being up to date all day at any location. It is desirable to consume news not only in front of the TV-set in the living-room, but anywhere in the house or even on the way to work, where the use of a mobile device, e.g. PDA or TabletPC, is much more comfortable. And why not combine the traditional *linear watching* of TV with the *interactive non-linear browsing* of the Internet on a personalized basis? Why not adapt the service to the – possibly time- and location-dependent – interests of the user?

This type of service, which is adaptable to different user needs, usable on different devices and which delivers content over different networks, needs two types of scaling: *content scaling* to adapt the media content to the specific requirements of the supported devices and *service scaling* to adapt the service to the specific requirements and properties of different device types, delivery mechanisms and user requirements. This includes delivering appropriately-scaled content and adapting to the device's modality (i.e. rendering and interaction capabilities).

Realizing scalable services requires knowledge of service components, device characterristics, the network and actual user needs. This information is modeled in a metadata set, the *service description*, which is described in the next section.

A Metadata Model Supporting Service Scalability

This section proposes a metadata framework for describing scalable interactive TV services. The model, outlined in figure 1, describes such a service as an enriched TV program. The model is detailed by introducing the underlying rationale that led to its development.

***Program* as anchor to traditional TV.** In the broadcast world, a program represents content linearly arranged along a time line. To comply with this traditional usage, it is essential that a metadata model does not change, but rather extends the notion of a linear program. The *Program* element at the root of the model

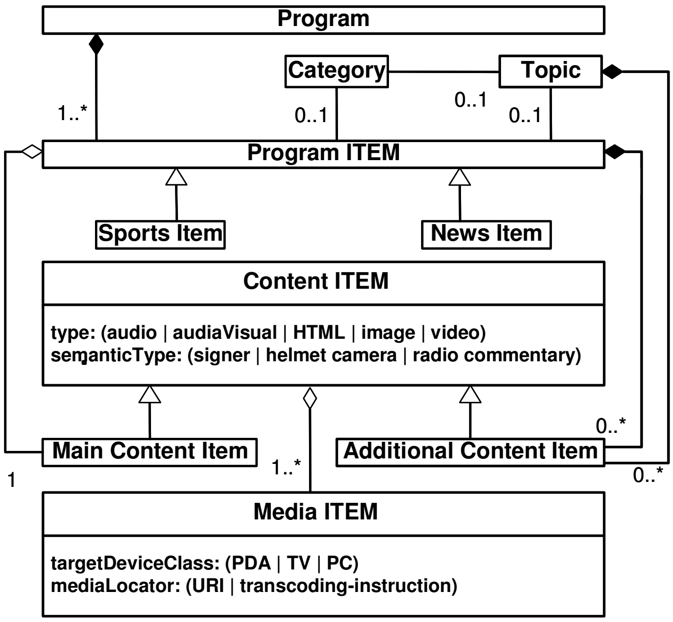

Figure 1: The SAVANT metadata model

[1] SAVANT, IST-2002-34814, is a project funded by the European Commission. See [2] for further information.

acts as the anchor to the traditional TV world, while it also serves as the basis for the proposed extensions.

Segmentation into *Program Items*. A program is represented by a sequence of *Program Item* elements, segmenting the linear TV content along the timeline. Each *Program Item* is considered as a coherent and semantically closed entity, such as a single news story (with its main and additional content). *Program Items* act as containers for both descriptive elements, including title, keywords and *Category*, and structural elements like the media content associated with a *Program Item*.

Device-independent content descriptions. The media content of a *Program Item* contains one segment of the linear broadcast content combined with a rich and diverse, but optional, set of additional content. To support the rendering of such content on various terminal devices, multiple device-specific formats of the same content may be included within the service. In order to allow describing the *same* content for various devices just once, the notion of *Content Items* is introduced in the model. *Content Items* provide a device-independent abstraction layer for describing only the *content* of the media without being specific regarding its concrete physical representation. At this level, those aspects are considered that influence how users interact with the content, such as its *type* (e.g. HTML page, audio track, video) and *semanticType* (e.g. signer, helmet camera), while the exact media format (e.g. MPEG-2, MPEG-4), is left open. Two types of *Content Items* are distinguished: *Main Content Items* (*MCI*) and *Additional Content Items* (*ACI*). *MCI*s describe the linear broadcast segments that make up the basic service. *ACI*s describe the supplemental information. For example, a news story in a News show, which is enriched with a signer and supplemented with background information from the Web, would be described as a *Program Item* with three *Content Items*: one *MCI* and two ACIs (the signer as an ACI of type "video" and the web page as an *ACI* of type "HTML"). *ACI*s may be marked as *synchronized* with an *MCI* by providing a reference to the timeline of the *MCI* in the metadata.

Device-specific media format. *Content Items* do not describe how the content is rendered at the terminal device. For instance, the signer *ACI* in the above example only specifies that it is a video clip to be synchronized with the main video (i.e. the *MCI*), but says nothing about its exact format. In the model, each concrete realization of a *Content Item* is described as a *Media Item*. These correspond to separate copies of the same content intended for use on different devices, each encoded in different formats or with different parameters (e.g. resolution, bit rate). A *Media Item* has a *targetDeviceClass* describing the device class for which the content was designed (e.g. TV, PDA), mapping directly onto basic technical parameters such as resolution or bit rate. Each *Media Item* contains a *mediaLocator*, which is a URI (Unified Resource Identifier) that either points to an actual, already existing essence or contains a transcoding instruction (the execution of which results in the creation of the appropriate essence). In order to render a *Content Item* at a terminal device, the *Media Item* that best matches the terminal type is chosen, and the appropriate player (e.g. MPEG-4 player or HTML browser) is automatically launched at the terminal.

Supporting non-linear access: Besides supporting the inclusion of *ACIs* at the *Program Item* level, it is also desirable to supply non-linear access to *Program Items* as well as to a group of several related *Program Items* (e.g. "Presidential election" in a news program or "Athletics world cup 2003" in a sports program). This latter type of access is provided by the *Topic* element. A *Topic* provides a "bigger story“. It can be described and classified; and additional content may be provided related to *Topics*. Typically, a *Topic* has a lifetime of a few days to a few weeks, during which new *Program Items* and new additional content may be added.

Implementing the Metadata Model

For the SAVANT system this metadata model has been mapped to a concrete metadata set, the *service description*, using a metadata language based on existing standards established or suitable for broadcast applications. The investigation of three standards, MPEG-7 [3], MPEG-21 [4] and TV-Anytime [5], led to the result that no single metadata language provides all the description elements necessary for describing the intended type of scalable TV services. Therefore, the service description has been realized by combining elements from all three standards. TV-Anytime is used to provide the framework, as its structure is best suited to describe the semantics of a TV service. The extensibility mechanism of TV-Anytime (replacing elements by those of derived types and including additional nodes into these types) is exploited to provide the additional elements as needed. Appropriate MPEG-7 elements were integrated, e.g., for describing media formats or summarizations. MPEG-21 elements, e.g. for device profiles, can also be included. This way, the extended description remains conformant with TV-Anytime.

A Content Access System Architecture supporting Service Scalability

The overall system architecture of SAVANT provides the end-to-end realization of producing, delivering and using the enriched interactive TV content. The *Content Access System* (*CAS*) presents the scalable service to the user by means of multiple heterogeneous devices. It is designed as a *Home Media Server* at the premises of the customer, which adapts the service such that it can be consumed in a personalized way using (currently) three different device classes with different properties: a conventional TV set for traditional viewing, a TabletPC as a portable powerful, highly interactive personal device and a PDA as a portable lightweight personal device. While the TV set is connected directly to the *Home Media Server*, the portable devices communicate with the server via a wireless connection.

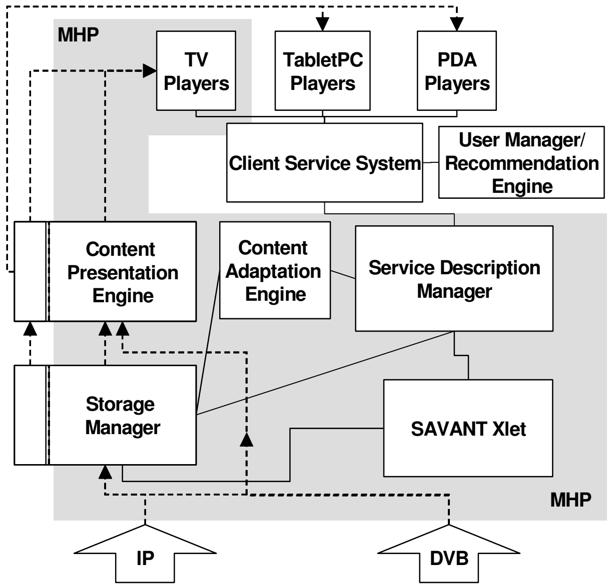

Figure 2: Content Access System Architecture

Figure 2 shows the main components of the *CAS* architecture, which is built upon MHP (Multimedia Home Platform) middleware. All service functionality is communicated to the *CAS* via the *service description.* The *Service Description Manager* (*SDM*) receives, updates, maintains and interprets the *service description*, and triggers processes depending on the information found there.

In order to realize scalable services, transcoding of content may be necessary in the *CAS*. The *SDM* instructs the *Storage Manager* to extract and store content. In the case that not all *Media Items* of a *Content Item* are simulcasted (i.e. provided in all the required formats), the *SDM* initiates transcoding using the *Content Adaptation Engine* in order to suit the requested device class.

The *SDM* also sends a copy of the *service description* to the *Client Service System*, which is responsible for the communication between the *Home Media Server* and the various client devices. Depending on the device characteristics, user interfaces with different functionality are provided. A *User Manager* and a *Recommendation Engine* are responsible for personalization, i.e. modifying the service and recommending service components based on the user's interests and preferences maintained in user profiles. User profiles, which are a record of a user's interests and preferences, are maintained. Such profiles can be built automatically by monitoring user actions over a period of time.

Finally, the requested content is delivered by the *Content Presentation Engine*, which also realizes the necessary presentation synchronization in cooperation with the *Service Delivery System* (*SDS*) at the service provider's premises. In the next section, we look in more detail at two specific functions of the *SDS*: smart routing and synchronization.

DELIVERY OVER CO-OPERATING BROADCAST AND BROADBAND NETWORKS

The Quest for Network Convergence

In the broadcast environment, the basic service components will always be delivered on the DVB channel for economical reasons as well as to meet the quality requirements of the consumer and service provider and to be backwards compatible with traditional services.

For multimedia-enriched services, where service providers can use the *ACIs* to exploit their content repositories in a new way and to distinguish themselves from other providers, new flexible ways of transmitting content must be found. An *ACI* can be synchronous to the main content (e.g. a signer translating news into sign language or a second camera view in a sports program), or asynchronous (e.g. background information like an audio clip or HTML pages, which can be consumed independent from the main content). In addition, the demand for enriched content may vary: low demand can be expected for content valuable only for a minority of consumers like the signer for hearing impaired people, and high demand is likely for popular content like a backstage video of a show.

Considering the available bandwidth and the delivery cost for the DVB channel, the integrated usage of broadband technology comes into focus. A preferable solution for broadcasters would be the flexible selection of the delivery mechanism dependent on actual cost and number of users. Additionally, switching to another channel should be possible to react to changes in the adopted cost model. The following sections describe mechanisms SAVANT provides to fulfill these requirements.

Smart Routing of Additional Content

Smart routing in the SAVANT system means selecting the most appropriate transmission mechanism for the distribution of each *Media Item* of a service, thus ensuring efficient and economical use of the mechanisms available. In the combined DVB/DSL environment, the starting point is to always transmit the main TV content via DVB. For the *ACI*s, several transmission protocols, on top of DVB and DSL, are available as alternatives in SAVANT's *SDS* (see figure 3).

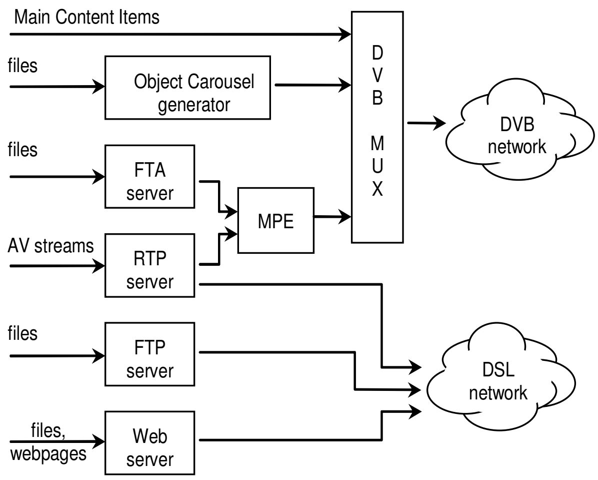

Figure 3: Available transmission mechanisms in the Service Delivery System (SDS)

The *Media Items* are contained in MXF containers (not shown). From these containers, the main content is played out directly via DVB transport streams (TS). Additional files can be transmitted via the DVB Object Carousel, the

File Transfer Application protocol[2] (using Multi Protocol Encapsulation, MPE), or made available via FTP or HTTP on DSL. Additional audiovisual[3] (AV) streams are transferred using RTP, encapsulated in IP, either on DVB (using MPE) or on DSL.

The most suitable transmission mechanism (or "*route*") can be determined by considering the following (at some points interrelating) dimensions:

- *content type* – The timing requirements of *ACI*s vary: whereas *synchronous* (streamed) content has strict timing[4] constraints, the timing of *asynchronous* content is not critical.
- *flexibility* - Routing flexibility defines whether or not the transmission mechanism for an *ACI* can be changed by the system. *Fixed routing* means the delivery mechanism is determined prior to starting the playout of an *ACI* and that it can not be changed afterwards. *Re-routing* means that the route can be changed while an *ACI* is playing.
- *characteristics of transmission mechanism* - For transmitting an *ACI*, one mechanism will be more suitable than another. Requirements of the *ACI* (e.g. timing constraints) must match the transmission mechanism's properties (e.g. real-time capability).
- *decision basis* - As a general rule, *ACI*s consumed by many viewers should be routed via DVB. The final routing decision may be taken based on metadata (*pre-defined*, e.g., by the provider), on popularity *predictions*, on popularity *measurements*, or on *operator decisions* at play-out time.

Based on these dimensions, SAVANT distinguishes four scenarios for content routing:

1. *Fixed routing of asynchronous ACIs*: asynchronous *ACI*s are inserted into either the DVB or DSL channel depending on a predefined field in the *service description*.
2. *Re-routing of asynchronous ACIs*: if the number of simultaneous users of an AV *ACI* accessed over DSL exceeds a predefined threshold, the *ACI* will be inserted into the DVB stream using the FTA protocol. If the usage figures drop again, the ACI is no longer made available via DVB but can still be pulled via DSL.
3. *Fixed routing of synchronous ACIs*: synchronous *ACI*s are inserted into either the DVB or DSL channel depending on a predefined field in the *service description*.
4. *Re-routing of synchronous ACIs*: if the number of simultaneous users of an AV *ACI* accessed over DSL exceeds a threshold, the *ACI* will be inserted into the DVB stream using the MPE protocol[5].

Synchronizing Additional Content with the Main Content

SAVANT has implemented a mechanism that allows for end-to-end synchronization of streamed *ACIs* with the main content, see figure 4.

For each audiovisual stream, an accompanying time-base is generated in the *SDS* that consists of timestamps transmitted in parallel with the content. The timestamps for the main content consist of NPT (Normal Play Time) descriptors, which are inserted into the MPEG-2

---

[2] The File Transfer Application protocol (FTA) is a proprietary protocol specifically developed to allow transfer of files to closed user groups in a broadcast network.

[3] The term audiovisual here is used to designate a piece of content that contains audio, video or both.

[4] Note that "strict synchronization" does not mean "lip sync" but the degree of synchronization required by synchronizing, e.g., a signer with the main program.

[5] Due to the number of issues associated with seamless switching and network congestion, this scenario will not be pursued by the SAVANT project.

TS [6]. The timestamps for streamed ACIs (transmitted via RTP, either on DVB or on DSL) are carried as 32-bit values in the timestamp field of the RTP packets [7].

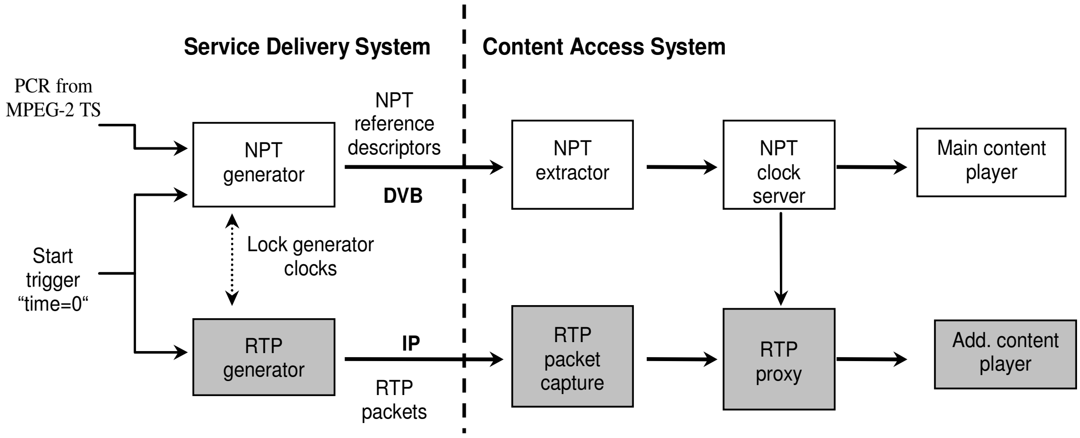

Figure 4: End-to-end synchronization of streamed additional content with main content

The NPT value represents a reference to the system time clock (STC) value of the TS. The timestamp in an RTP data packet reflects the sampling time of the frame to which the first octet in that packet contributes. The generation of the NPT descriptors and RTP timestamps is triggered to start simultaneously. Their value is set to "0" at the start of the main program and increased periodically [6]. In this way, a common time base for MCIs and ACIs during the TV program is realized.

In the *Content Presentation Engine* of the *CAS*, the extracted NPT values are used to generate a local clock, which is managed by the *NPT Clock Server*. The *RTP Proxy* presents an RTP packet to the *ACI Player* when the local clock in the *CAS* reaches the value of the timestamp in that packet. Thus, the *CAS* can delay the RTP stream if needed, and resynchronize it with the MPEG2-TS (i.e. the main program).

A SERVICE EXAMPLE: ENHANCED NEWS

Based on the concepts described above, the SAVANT project has implemented a system that supports scalable services and is able to deliver service components via DVB and/or DSL co-operatively. This system is demonstrated by means of an enhanced News service.

The service consists of a series of TV News programs, each comprising of a series of *News Items*. Supplementing the main news content, related *ACIs* are transmitted, which provide background information in the form of HTML pages, audio and video clips. Each of these items is either associated with an individual *News Item*, or with a *Topic*. Additionally to these *Asynchronous ACIs*, a signer video supporting hearing-impaired viewers and subtitles in a foreign language are transmitted *synchronously* to the main program and presented if the profile of the currently watching user requests them.

The *Content Access System* records the news broadcast. This way, both main and asynchronous additional content can be accessed either "live" or at a later time. Asynchronous additional content not available at the time of viewing is pulled, transparently to the user, by the *CAS* from the *SDS* via the DSL channel. Access to synchronous additional content is restricted to "live" viewing. This content is sent using smart routing via the currently most suitable transmission mechanism, again transparently to the user. The service is available on three different device classes: the TV set, the Tablet PC and the PDA. On the Tablet PC and the PDA, no additional software is needed besides a Web

---

[6] Ideally, the timestamp generator clocks should be locked in order to prevent them from drifting.

browser and appropriate media players. The user interface is designed in a way that the user needs only the remote control on the TV set or a stylus on the other devices. For browsing through the *News Items* by program, *Topic* or *Category*, the available items are listed. A simple click launches the presentation of the selected item. Users can also compose their own personalized program by either manually selecting *News Items* of interest, or by relying on recommendations based on a dynamic personal user profile.

CONCLUSIONS

In this paper, a concept of scalable interactive TV services has been proposed, tailored to be consumed using a variety of diverse terminals. A metadata schema has been introduced describing all components making up a scalable service, their relationships and properties.

Furthermore, the SAVANT system infrastructure has been described, which allows transmitting such services via a combination of DVB and broadband IP networks. The main program is always sent via the broadcast network. Additional related content is transmitted either via the broadcast channel for a large number of viewers or via a broadband IP channel for smaller audiences. For each individual *Media Item*, the proposed system selects the most suitable transmission mechanism automatically, depending on content parameters, network state and number of viewers.

The system supports the end-to-end synchronization of media streams even if they are transmitted over different networks in order to ensure that the viewer experiences a seamless service.

As an example, an enhanced interactive News broadcast has been described. Although this is a specific service, the system is open to all genres of Rich Media interactive TV.

REFERENCES

1. Y. Gonno et al., “White paper on integrated broadband environment for personalized TV experience (IBEX),” in *Proc. ACM Multimedia*, 2000, pp. 63–66.
2. SAVANT project home page, `http://www.savant.tv`.
3. ISO MPEG-7, Part 5 - Multimedia Description Schemes, ISO/IEC JTC1/SC29/WG11/N4242, 2001.
4. ISO MPEG-21, Part 7 - Digital Item Adaptation, ISO/IEC JTC1/SC29/WG11/N5231, 2002.
5. ETSI TS 102 822-3-1: Broadcast and On-line Services: Search, select and rightful use of content on personal storage systems ("TV-Anytime Phase 1"), Part 3 Metadata, Sub-part 1: Metadata Schemas. 2002.
6. ISO/IEC. Information Technology - Generic coding of moving pictures and associated audio information - Part 6: Extensions for DSM-CC. Geneva, 1998. (ISO/IEC 13818-6)
7. IETF Network Working Group. Request for Comments: 3550 - RTP: A Transport Protocol for Real-Time Applications, July 2003 (http://www.ietf.org/rfc/rfc3550.txt)

ACKNOWLEDGEMENTS

The work presented is part of the SAVANT project funded by the European Commission under contract number IST-2001-34814. We thank all members of the SAVANT consortium, especially Gabriella Kazai, Gunn-Kristin Klungsøyr, Michael Probst and Michael Niedermoser. This publication derived many benefits from for their contributions to the concept, design and realization of SAVANT.