ForgetIT

ForgetIT Project, GA 600826

ForgetIT Project Presentation

ForgetIT Consortium

- ForgetIT Information Package (D11.1)
- April 2013
-

A Computer that forgets ?
Intentionally ??
And in context of preservation???

- **However** we are facing
  - dramatic increase in content creation (e.g. digital photography)
  - information overload and changing professional + private lives
  - increasing storage costs for long-term storage (>10 year)
  - increasing use of mobile devices with restricted capacity
  - inadvertent forgetting in lack of systematic preservation

- **And:** Forgetting plays a crucial role for human remembering and life in general (focus, stress on important information, forgetting of details)

So: Shouldn't there be something like forgetting in digital memories as well? → ForgetIT

Motivation

ForgetIT

Needs

- increasing amount of digital content handled over decades
- more or less systematic backup strategies used
- non-paper practices for long-term perspective required

Major Obstacles

- large gap for adoption
- high-up front cost
- no established practices
- lack of understanding of benefit
- reluctance to invest

Opportunities

- major progress in preservation technology
- maturing Information extraction technology
- storage as service (e.g. clouds)

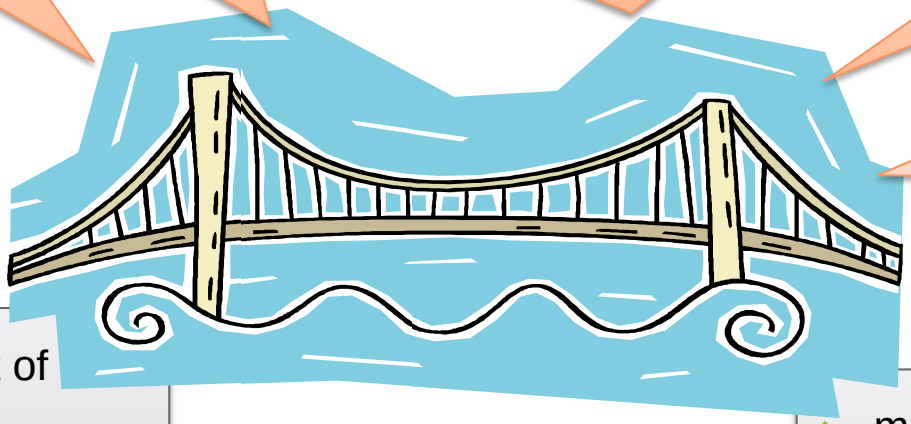
Enabling smooth transition to preservation

Creating immediate benefit + reducing effort

Opening alternatives to "keep it all" and "forgetting by accident"

Easing interpretation in the long run

taking inspiration from and complementing human memory



Needs

- increasing amount of digital content handled over decades
- more or less systematic backup strategies used
- non-paper practices for long-term perspective required

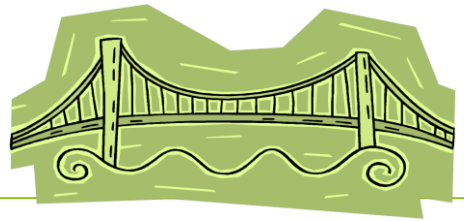
Major Obstacles

- large gap for adoption
- high-up front cost
- no established practices
- lack of understanding of benefit
- reluctance to invest

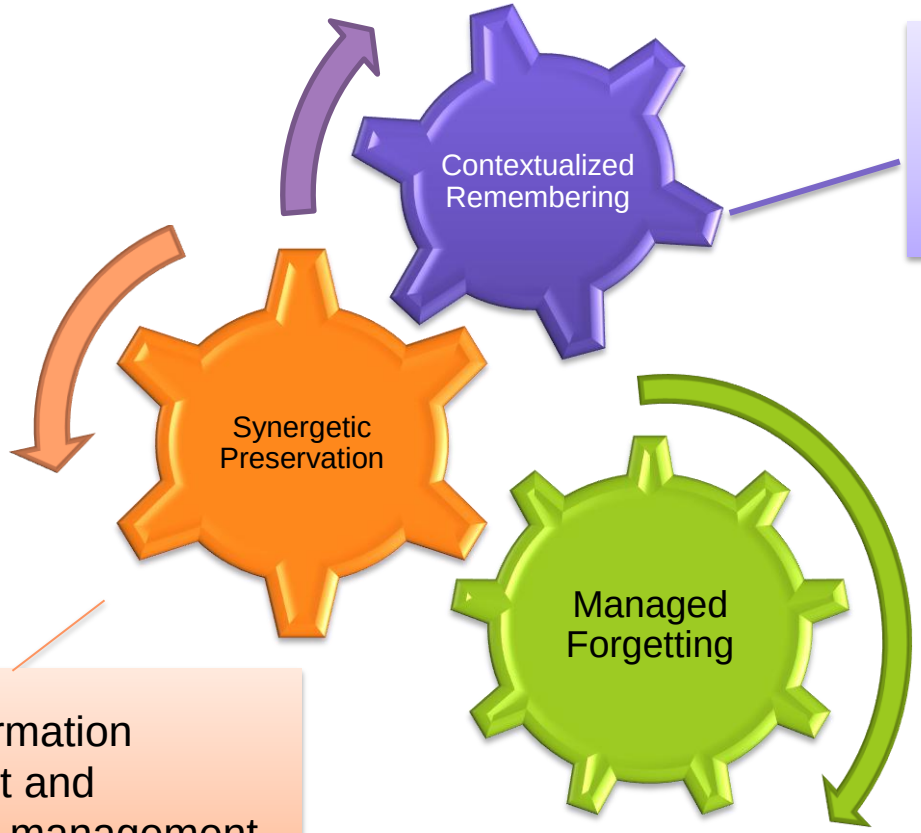
Opportunities

- major progress in preservation technology
- maturing Information extraction technology
- storage as service (e.g. clouds)

Building the Bridge



ForgetIT



- bringing back information into active use in a meaningful way

- couples information management and preservation management

- as opposed to the current "forgetting by accident"
- inspired by human forgetting

Managed Forgetting

ForgetIT

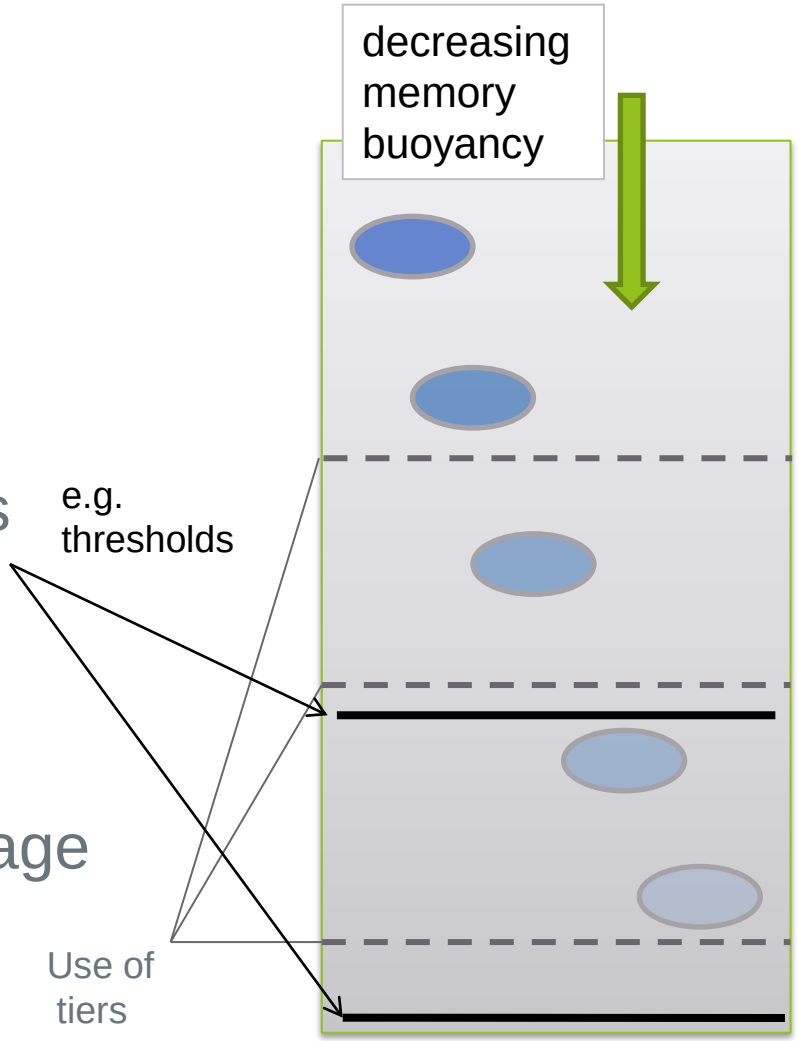
- Aim:
  - help in identifying and focus on relevant information
  - supporting preservation content selection
- inspired by central role of human forgetting
- managed forgetting will replace inadvertent forgetting
- managed forgetting ≠ automatic deletion
- instead: range of forgetting options e.g.
  - resource condensation
  - influence of indexing
  - reduction of redundancy

Managed Forgetting

ForgetIT

… is based on:

- careful information value assessment
  - Memory buoyancy
  - Preservation value
- forgetting strategies via policies (control risk level)
- option to integrate final manual checking before deletion
- combination with multi-tier storage solution possible



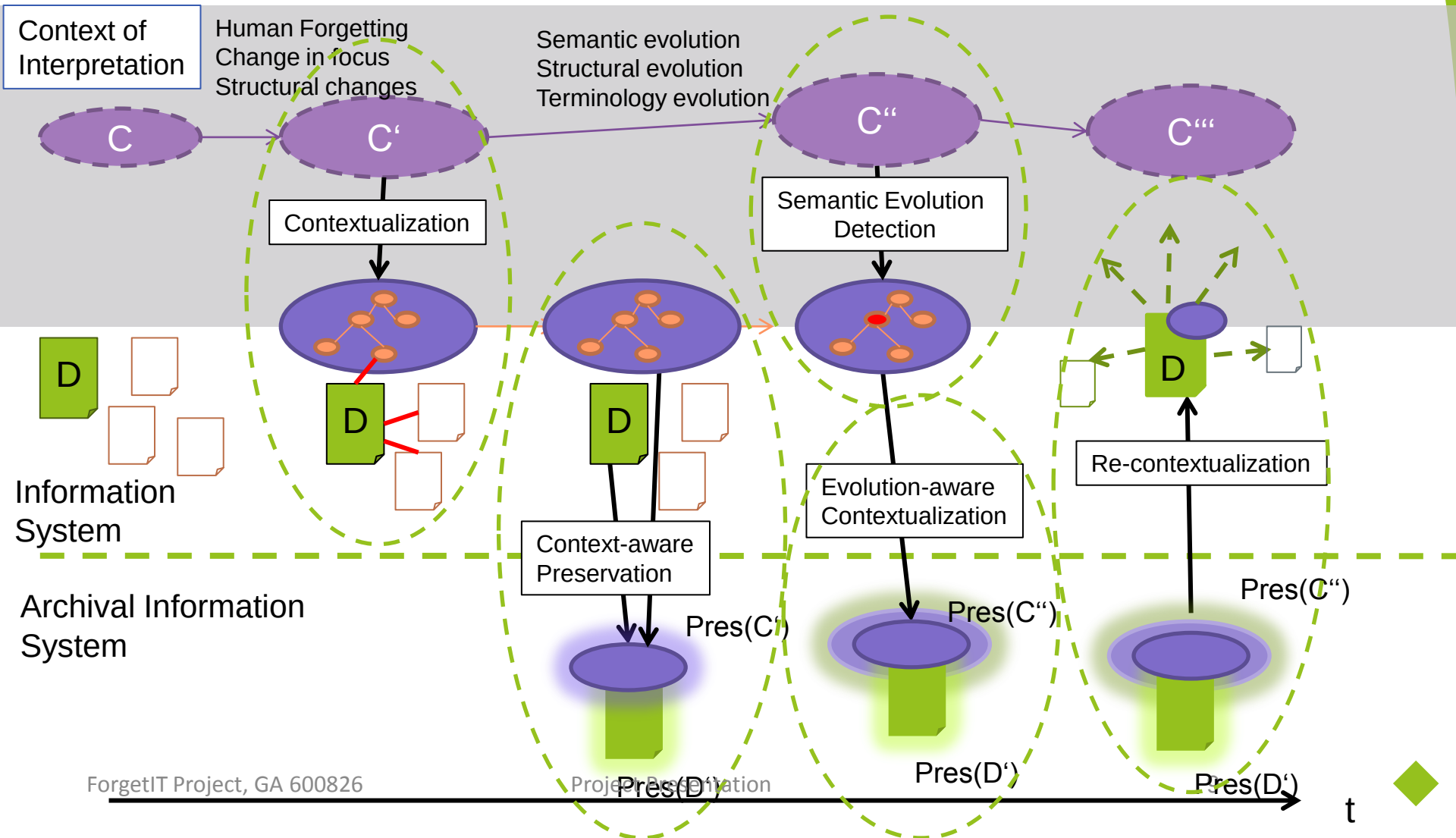
Contextualized Remembering

ForgetIT

- Aim:
  - bringing back information into active use in a meaningful way even if a lot of time has passed
  - aiming for semantic level of preservation
- taking into account relevant parts of context when moving to archive
- increasing contextualization of preserved content
- considering context evolution over time (evolution-aware contextualization)
- aiming for semantic level of preservation

Evolution-aware Contextualization & Re-contextualization

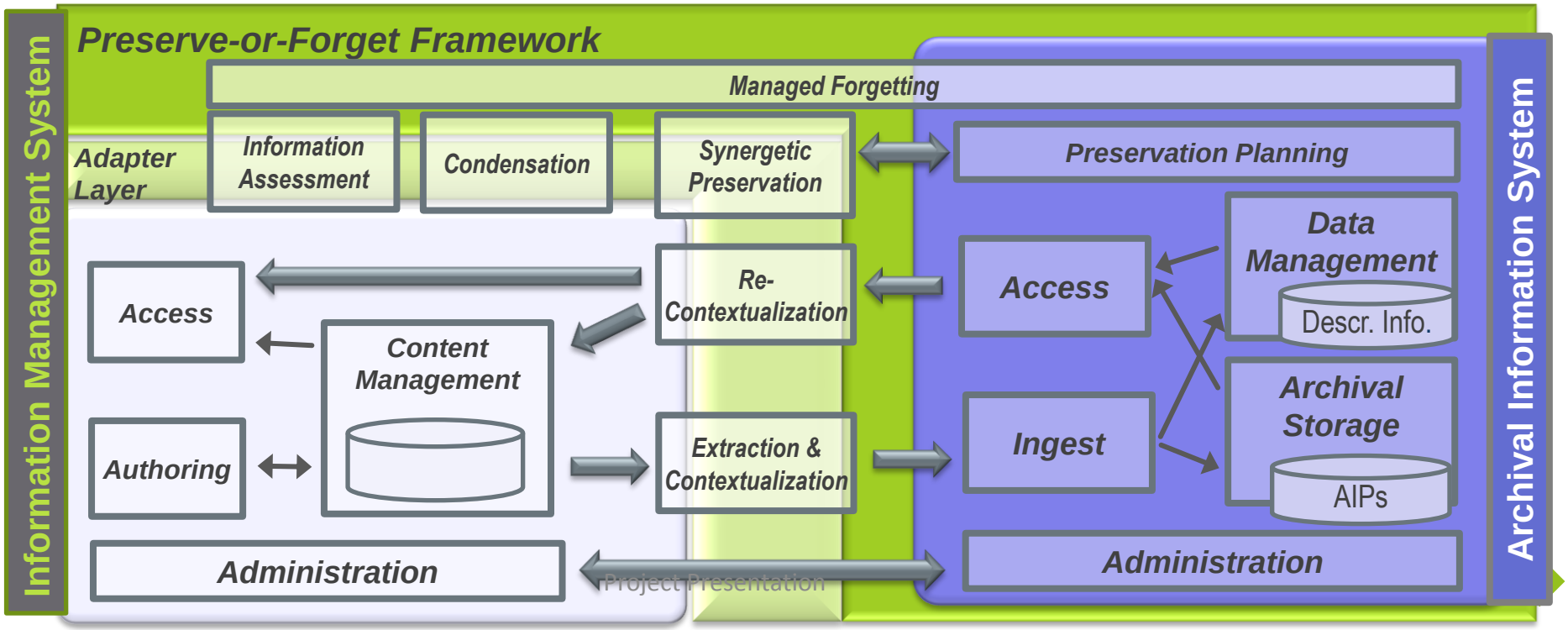
ForgetIT



Synergetic Preservation

ForgetIT

- Aim: smooth and step-wise transition between active information use and preservation
- enables rich information flow in both directions
- supports more informed preservation decisions
- eases preservation adoption



ForgetIT Pilot: Organizational Preservation

ForgetIT

Starting point: existing and popular CMS (TYPO3)

- Sophisticated worklflows for content creation and publication
- **But:** Separation of publication and preservation/archival

→ Acssess to archived content is difficult and costly

→ obsolete and even outdated information stays online

ForgetIT approach:

- Preservation as integral part (binary model → gradual managed forgetting)
- Bolder attitide towards removing content from active use
- Automated support of cleaning up processes
- Support of many stages of archiving, e.g. offline but still in index, aggregates online/ content in archive, only aggregates kept, etc.

Dissemination/Exploitation:

TYPO3 with preservation extension as open source project to TYPO3 community (involvement of TYPO3 Community)

ForgetIT Pilot: Personal Preservation

ForgetIT

Starting point:

- Tremendous growth of information in personal sphere
- Diversity and fast evolution of devices, platforms and formats
- **Keeping info sustainably available:** Ad hoc solutions for mid-term, long-term solutions missing (esp. private)

ForgetIT approach:

- Preservation solution for personal information space
- Based on concept of Semantic Desktop
- Consideration of social web content, multimedia content, other types of personal content, knowledg structures
- Additional short/mid-term benefit: de-cluttering information space by managed forgetting
- Consideration of multi-level infrastructures (e.g. mobile, PC, cloud)

Dissemination/Exploitation: Personal Preservation as a service

The ForgetIT Consortium

ForgetIT

- Leibniz Universität Hannover - L3S Research Center (Coordinator)
- Luleå University of Technology
- IBM Israel Science & Technology Ltd. (IBM Research – Haifa)
- Türk Telekom
- Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI)
- Centre for Research and Technology Hellas (CERTH)
- dkd Internet Service GmbH
- University of Sheffield
- University of Edinburgh
- EURIX Group
- University of Oxford

RTD Activities in ForgetIT

ForgetIT

Research Organized in 4 Research Areas (RAs):

- The Forgetting Process (RA-1)
- Information Condensation & Consolidation for Managed Forgetting (RA-2)
- Information Contextualization and De-Contextualization for Preservation (RA-3)
- Synergetic Preservation – Joint Information & Preservation Management (RA-4)

The Forgetting Process (RA-1)

ForgetIT

- Interdisciplinary foundations: Forgetting and remembering in human and digital memory
- Assessing human expectations towards digital forgetting
- Multifaceted information assessment in support of managed forgetting
  - Factors for Memory Buoyancy and Preservation Value
- Implementation of the Forgetting process
- Options and policies for preservation and managed forgetting

Information Condensation & Consolidation for Managed Forgetting (RA-2)

ForgetIT

- Consideration of textual and multimedia content
- Analysis for content similarity and redundancy
- Semantic analysis for condensation
- Information condensation and consolidation based on usage, diversity and coverage

Information Contextualization and De-Contextualization (RA-3)

ForgetIT

- Models and approaches for information contextualization for Preservation
- Knowledge extraction methods for relevant context dimensions
- Methods for content de-contextualization for packaging content into independent preservation packages
- Methods for dealing with evolving semantics and evolution-aware contextualization
- Methods for Re-contextualization in support of contextualized remembering

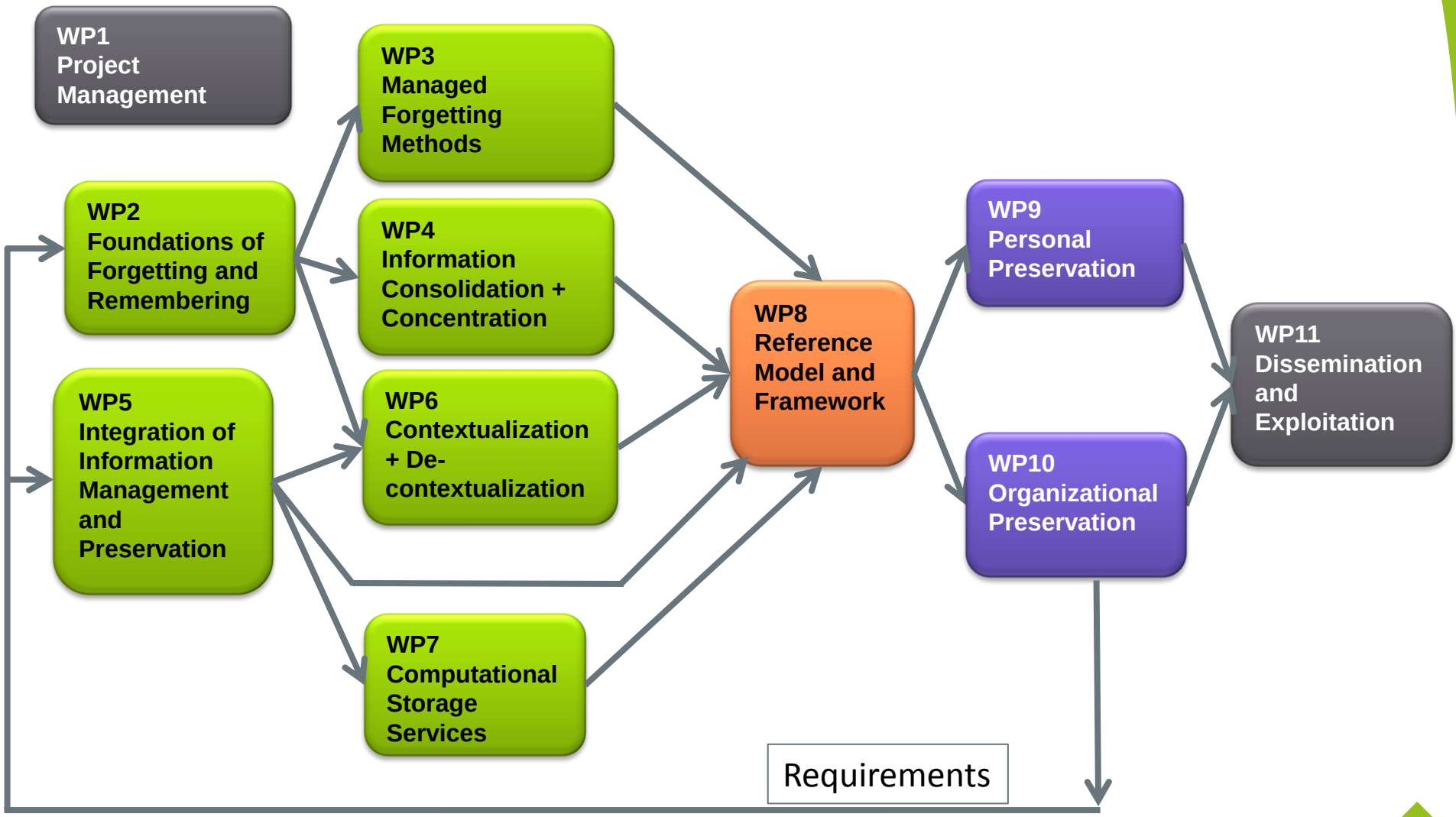
Synergetic Preservation – Joint Information & Preservation Management (RA-4)

ForgetIT

- Work on conceptual level:
  - Preservation reference model extensions
  - Integration of preservation in content management workflow
- Design of an Integration architecture, which enables synergetic preservation
  - Coupling of information and preservation management system
  - Integration of managed forgetting process
  - Integration of contextualized remembering
- Models and methods for smooth bidirectional transitions between information management and preservation storage
- Computational preservation storage

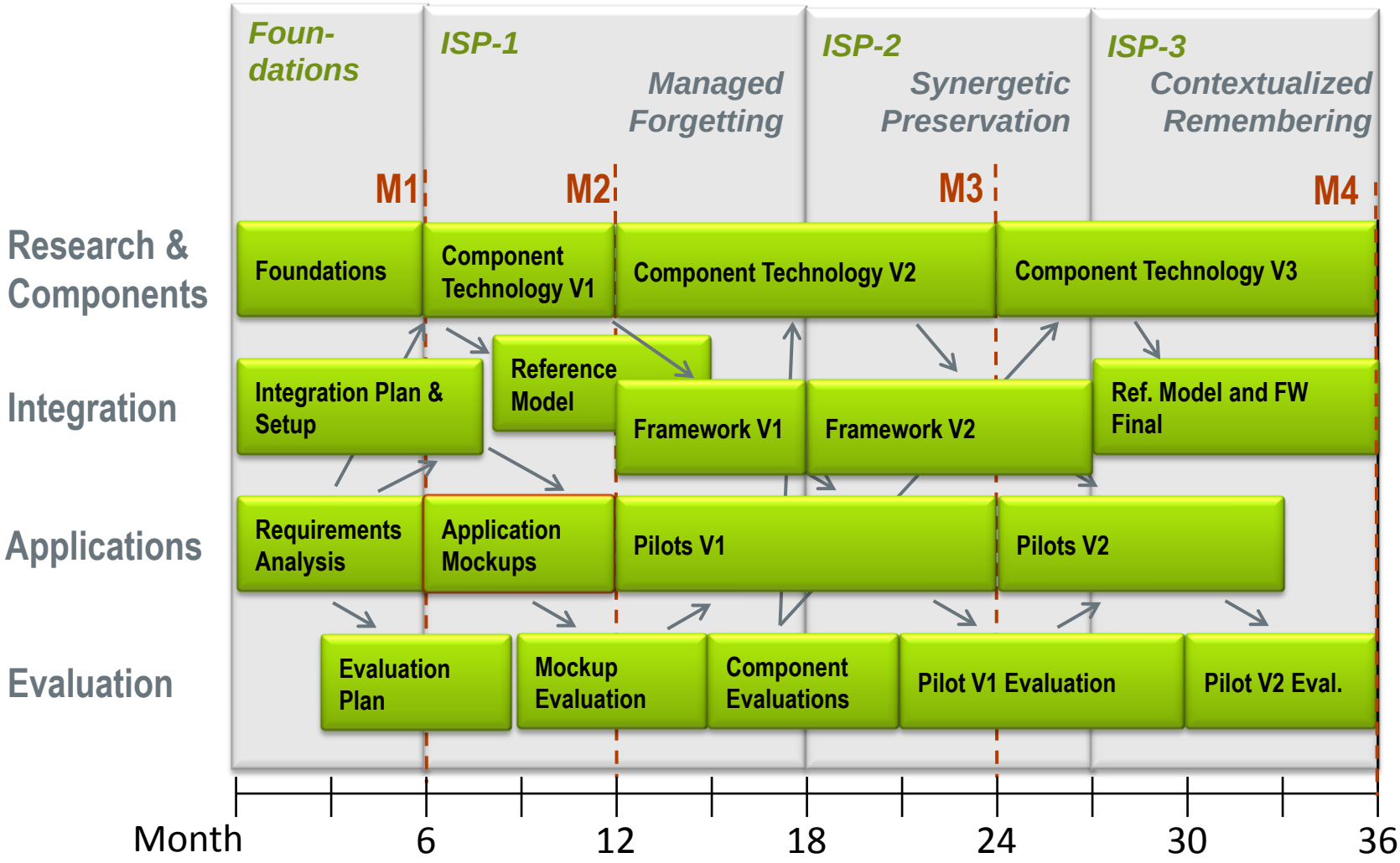
Project Structure: Work Packages

ForgetIT



Project Schedule Overview

ForgetIT



Expected Outcome

ForgetIT

- Foundations & Models:
  - Approaches for managed forgetting and contextualized remembering
  - Joint content lifecycle model for synergetic preservation
- Algorithms & Methods for:
  - preservation oriented summarization and aggregation
  - multifaceted information assessment methods
  - evolution-aware contextualization and re-contextualization
  - storage based computation
- Infrastructure & Services:
  - Preserve-or-Forget Framework
- Application Pilots for
  - personal preservation focusing on multimedia coverage of personal events
  - organizational preservation focusing on smooth preservation in organizational content management
- Best practices and adoption blueprints