arcomem

EXPLOITING THE SOCIAL & SEMANTIC WEB
FOR GUIDED WEB ARCHIVING

ARCOMEM's aim is to help *transform archives into collective memories* that are more tightly integrated with their community of users and *exploit Social Web and the wisdom of crowds* to make Web archiving a more selective and meaning-based process.

Motivation

The Web is a quickly changing, ever growing information space

- It's growing by >8% per week
- After 1 year only 40% of the pages are still accessible while 60% of the pages are new

! **A Web Archive as a Collective Memory is a cultural necessity for the future**

! **But „Archive and Store Everything" is not a practical approach**

Approach

Make use of the Social Web

- Huge source of user generated content
- Wide range of articulation methods From simple „I like it" Buttons to complete articles
- Represents the diversity of opinions of the public

User activities often triggered by

- Events and related entities (e.g. Sport Events, Celebrations, Crises, News Articles, Persons, Locations)
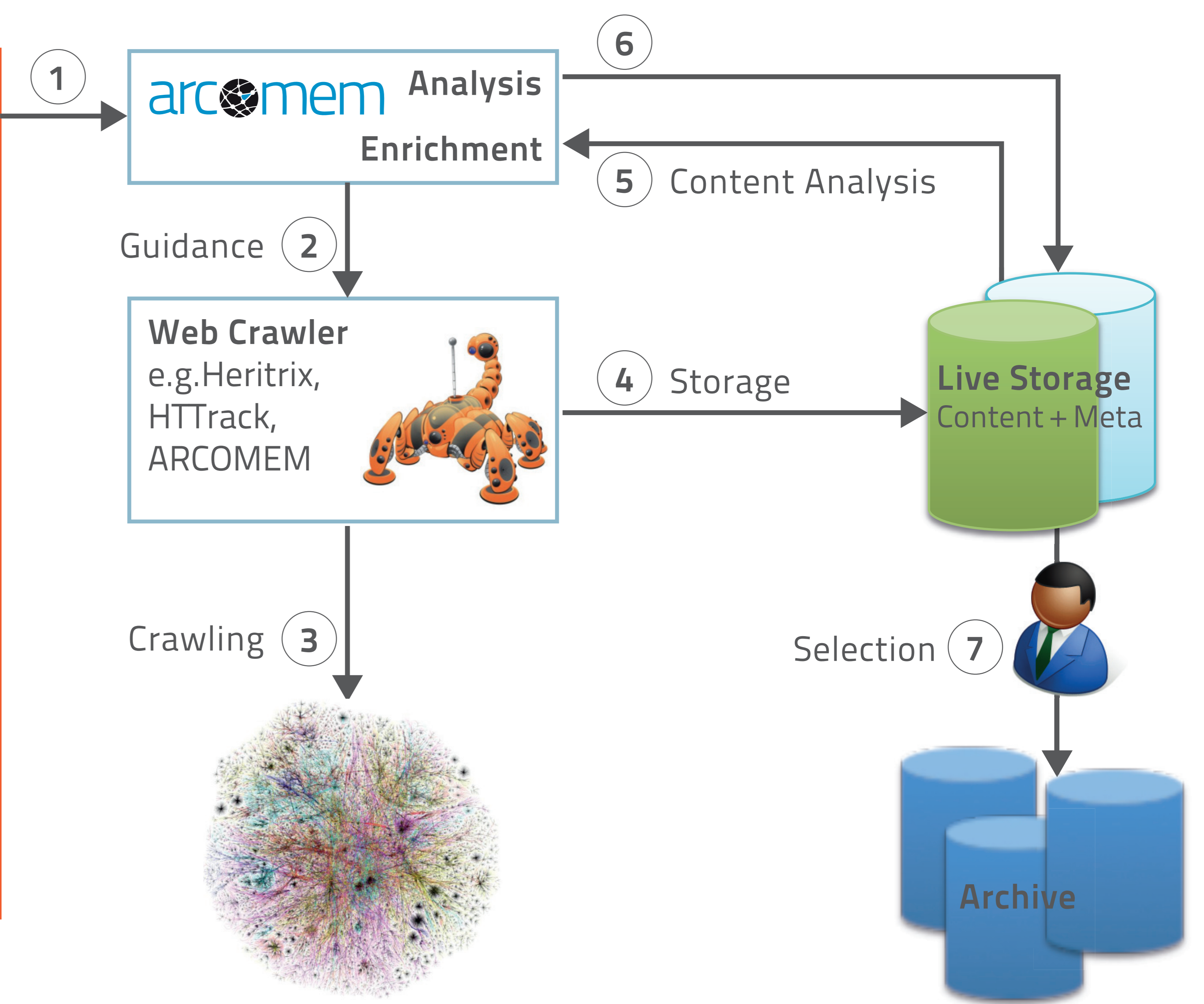- Topics (e.g. Global Warming, Financial Crisis, Swine Flu)

A semantic-aware and socially-driven preservation model is a natural way to go.

Entities, Events
Europe, Greece, Germany, Sarkozy, Merkel, Papandreou, …

Media Categories
Microblogs, Social Networks, …

Reference Seedlist
http://www.economist.com/node/21534849
http://www.ekathimerini.com/ekathi/comment
http://www.bbc.co.uk/news/world-europe-15589568

Seedlist
http://www.bbc.co.uk/search/news/?q=Greek%20crisis
http://www.guardian.co.uk/business/blog
http://www.kathimerini.gr/
http://twitter.com/#!/EU_Commission

1 → arcomem **Analysis Enrichment** 6

Guidance 2

5 Content Analysis

Web Crawler e.g.Heritrix, HTTrack, ARCOMEM

4 Storage → **Live Storage** Content + Meta

Crawling 3

Selection 7

Archive