



Aligned, Quality-centric Software and Data Engineering



ALIGNED will allow European data and software engineering industries to exploit new opportunities in web data by developing models, methods and tools for engineering information systems that use big data.

AT A GLANCE

Project title:

Aligned, Quality-centric Software and Data Engineering

Projects coordinator

Dr Rob Brennan, TCD

Partners from:

Trinity College Dublin (Ireland); University of Oxford (UK); Universität Leipzig (Germany); Semantic Web Company GmbH (Austria); Wolters Kluwer Deutschland GmbH (Germany, Poland); Adam Mickiewicz University, Poznan (Poland).

Duration: 3 years

Total cost: €4m

EC Contribution: €4m

Programme: Horizon 2020

Further information:

www.aligned-project.eu

Context and motivation

The rapid growth of web data creates demand for software engineering methods which can build and maintain applications that extract, process and publish this data. ALIGNED will develop new models, methods and tools for engineering information systems based on co-evolving software and web data.

Challenge

Currently, software and data engineering lifecycles are disconnected and separate. The tools, techniques, and lifecycles of both can easily fall out of sync, hindering innovation and making quality assurance difficult. These systems often require extensive customisation to reflect changes in user requirements, organisational policies, and the structure and interpretation of the data they hold. This customisation is expensive and error-prone. Large, complex software and data systems mean that exhaustive testing is necessary before any new feature can be added to the existing design. As software and data changes during their lifecycles, engineers are forced into a choice between expensive modification and continued operation with an inefficient design.

Solution

ALIGNED will create a combined software and data engineering meta-model that captures the data lifecycle, domain knowledge, and design intentions. This meta-model will allow users to describe their software and data engineering projects. From this meta-model, ALIGNED will create tools for software and data engineering of data-intensive systems. The software tools will use this model to produce software design models, transformations that generate or configure data-intensive applications. The data tools will use the meta-model to produce data development models including data quality and integrity constraints, data test frameworks, data curation workflows and data transformations. In addition, collaboration or process tools will help to integrate the outputs of both engineering processes. ALIGNED will develop methods for using these meta-models and tools as part of a software and data engineering process, with an emphasis upon techniques for ensuring data quality and integrity, as well as software security and reliability.

These tools and processes will improve the way in which users handle integrated data and software engineering. Users will be able to automatically detect mismatches between their software and data engineering efforts. Automated correctness checking tools will ensure that data remains compliant with schemas throughout its lifecycle. Automatic generation of tools and widgets will reduce the necessity of investing large amounts of time and energy in customisation. Mass participation tools will allow the generation and curation of large and complex high-quality datasets at lower costs in time and effort than previously. Expert-level tools will allow users to build analyses on top of these datasets, reconciling complex and conflicting data and transforming it into authoritative interpretations. Visualisation and display tools will allow for quick and easy publication of these datasets and analyses in comprehensible formats for non-expert users to consume and use.

Expected impact

ALIGNED will provide a comprehensive suite of tools and methodologies for combined software and data engineering. These tools and methodologies will reduce customisation costs and increase agility. Consumers in fields such as data-intensive application development and knowledge management platform provision will be able to use this suite to improve the integrity, robustness and reliability of their systems. ALIGNED will ensure the availability of innovative tools for handling complex software systems, which will provide a significant productivity increase in the development, testing, verification, deployment and maintenance of data-intensive systems and highly distributed applications.

ALIGNED will provide targeted impacts on each of our four use cases. For the Semantic Web Company's PoolParty product, ALIGNED will produce tools to manage schema complexity and assist software and data co-evolution. ALIGNED will help streamline the Wolters Kluwer JURION content pipeline, improve searching, and reduce errors in data transformation. The Seshat Global History Databank will gain tools to validate data, improve data collection, and for experts to aggregate data into analyses. ALIGNED will produce tools to assist communities with changing datasets and to map unstructured to structured data for DBpedia.

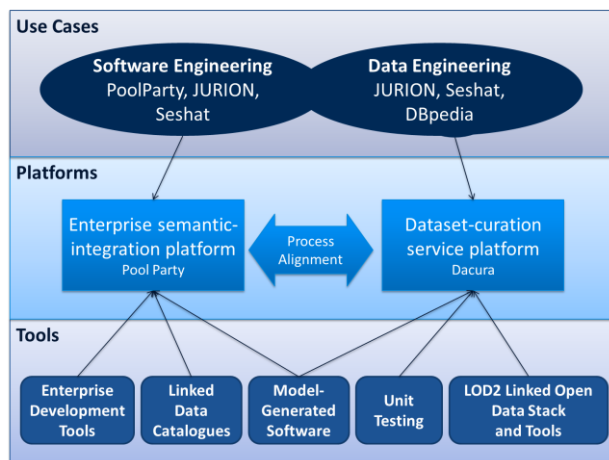


Figure1: ALIGNED use cases and tools