

AQUA – UM SISTEMA DE INFORMAÇÃO PARA ANÁLISE E VALIDAÇÃO DE PARÂMETROS DE QUALIDADE DA ÁGUA EM ALQUEVA

Orlando BELO¹; Anália LOURENÇO¹; Paula SARMENTO²; Ana MAGRIÇO²; José Luís PINHO³; Manuela LIMA³; José VIEIRA³

*¹Departamento de Informática, Escola de Engenharia, Universidade do Minho
Campus de Gualtar, 4710-057 Braga, PORTUGAL*

*²Empresa de Desenvolvimento e Infra-estruturas de Alqueva S.A.
7800-522, Beja, PORTUGAL*

*³Departamento de Engenharia Civil, Escola de Engenharia, Universidade do Minho
Campus de Gualtar, 4710-057 Braga, PORTUGAL*

RESUMO

A fiabilidade de dados disponíveis em domínios como os da hidrologia, das infra-estruturas (hidráulicas e sanitárias) e das variáveis biofísicas e sócio-económicas constitui elemento de base fundamental para a alimentação de qualquer sistema de suporte à decisão numa bacia hidrográfica. Isto significa que a transformação de dados brutos em informação com qualidade controlada, deve ser processada de forma eficiente e compatível com exigências dos ambientes hidroinformáticos, dos decisores e do público em geral.

O sistema de informação AQuA, criado no âmbito do projecto ODeAnA, tem como objectivo principal o desenvolvimento de um sistema para análise e validação de parâmetros de qualidade da água no Empreendimento de Fins Múltiplos de Alqueva. Este sistema visa recolher os dados provenientes de leituras sobre a qualidade da água efectuadas pelas estações de recolha automática ou por técnicos de laboratório, analisando e validando esses mesmos dados e, posteriormente, integrando os dados limpos e consolidados numa base de dados projectada especificamente para o efeito. O sistema assegura mecanismos de detecção de anomalias, bem como a manutenção de esquemas de resolução para alguns tipos dessas ocorrências. Em termos de análise, é permitida a exploração dos dados, tendo o apoio de algumas funcionalidades de geo-referenciação, e a geração de relatórios ou gráficos específicos. O sistema está preparado para acolher os diversos perfis de utilização definidos: administrador de sistemas, supervisor de resultados, utilizadores internos e externos e público em geral.

A implementação do sistema foi organizada e desenvolvida segundo dois módulos operacionais: (i) AQuA Wrapper – responsável por tratar os serviços relacionados com extracção, processamento e armazenamento dos dados provenientes das diversas estações de recolha; (ii) AQuA Web – responsável por tratar os serviços relacionados com validação e exploração dos dados recolhidos. Este módulo disponibiliza uma plataforma Web para acesso aos dados resultantes das leituras efectuadas sobre a qualidade da água nas diversas redes do sistema e constitui o principal elo de ligação dos utilizadores com o sistema de monitorização.

Palavras-chave: Sistemas de Monitorização da Qualidade da Água, Hidroinformática, Parâmetros de Qualidade da Água, Processamento Analítico de Dados, JAVA e PHP.



1 - INTRODUÇÃO

A água é um bem precioso. Este é um facto que todos os dias vem sendo reforçado, conforme se vai tendo consciência da situação em que se encontram muitos dos recursos hídricos distribuídos pelo mundo. Actualmente, procuram-se os meios e mecanismos necessários para preservação desses recursos, de modo a assegurar a sua disponibilidade e qualidade. A protecção do “património” água exige que sejam tomadas medidas efectivas no controlo do acesso a esses recursos, assim como no controlo da poluição e na forma do seu consumo. Em termos internacionais, em particular na Europa, existem algumas situações delicadas a ter em conta, nomeadamente: os riscos de poluição das águas superficiais, a sobre-exploração das águas subterrâneas, o estado actual de algumas zonas húmidas e o aumento significativo da área de terrenos irrigados no Sul da Europa. Os efeitos causados pela poluição dos recursos hídricos são sobejamente conhecidos, com algumas situações de disfunções ambientais graves em certas áreas geográficas. Tornou-se hoje imprescindível a existência de instrumentos legislativos eficazes que ajudem no complexo trabalho de planeamento e de gestão sustentada dos recursos hídricos.

Estabelecendo um quadro de acção comunitária no domínio da política da água, surgiu a Directiva 2000/60/CE do Parlamento Europeu e do Conselho, de 23 de Outubro de 2000 [DQA 2002], normalmente reconhecida como *Directiva Quadro da Água* (DQA)¹. Esta directiva é um ambicioso instrumento de actuação no domínio do planeamento e da gestão de recursos hídricos no espaço da União Europeia, estendendo o âmbito de aplicação das medidas de protecção da água a todas as suas águas. Os seus objectivos primordiais são alcançar o “bom estado” de todas as águas europeias até ao ano de 2015 e assegurar a utilização sustentável da água em toda a Europa. Teoricamente, a DQA está orientada para uma visão moderna de gestão da procura e de gestão integrada da água e do território, considerando que a água não é um mero produto comercial, mas um património que deve ser protegido, defendido e tratado como tal. Esta política procura contribuir para a protecção e melhoramento da qualidade do ambiente. Os recursos naturais devem ser geridos de forma a ter em consideração as inter-relações que existem entre a água, o solo, a fauna e a flora, de forma a evitar disfunções ecológicas.

2 - O EMPREENDIMENTO DE FINS MÚLTIPLOS DE ALQUEVA

O Empreendimento de Fins Múltiplos de Alqueva (EFMA)² data de 1957, altura em que foi criado o Plano de Rega do Alentejo. Porém, os trabalhos em Alqueva só se iniciaram em 1976. Prolongaram-se durante algum tempo até serem interrompidos, em 1978. Quase quinze anos depois, em 1993, os trabalhos recomeçaram, novamente em fase de estudo e avaliação. Foi então criada a Comissão Instaladora da Empresa do Alqueva, que relançou o processo com o lançamento dos concursos públicos internacionais necessários. Passados dois anos esta comissão foi substituída pela Empresa de Desenvolvimento e Infra-estruturas do Alqueva, S.A. (EDIA) que, hoje, gere o EFMA.

O EFMA abrange um grande número de concelhos do Alto e Baixo Alentejo. O empreendimento desenvolve-se a partir da barragem de Alqueva, instalada no rio Guadiana, imediatamente a jusante da confluência do rio Degebe e a montante da confluência do rio Ardila. Tem grande importância estratégica para as regiões referidas e para o próprio país. Os seus principais objectivos são a constituição de uma reserva estratégica de água e a garantia de seu abastecimento regular às

¹ <http://dqa.inag.pt/>

² <http://www.edia.pt>



populações. Assim sendo, torna-se vital garantir a qualidade da água armazenada na sua principal albufeira, a albufeira de Alqueva, a qual se estende por 83 km ao longo dos concelhos de Moura, Portel, Mourão, Reguengos de Monsaraz e Alandroal.

A EDIA, principal entidade responsável pelo projecto, controla sistematicamente a qualidade da água do empreendimento, utilizando um sistema de estações de medição automática. Estas estações, distribuídas ao longo da albufeira em pontos estratégicos, fornecem, em períodos de tempo regulares, diversas leituras dos parâmetros utilizados em processos de avaliação da qualidade da água. No sentido de otimizar o processamento dos valores dessas leituras, melhorar o controlo sobre as estações de recolha de dados, reagir de forma mais efectiva e adequada a eventuais falhas de controlo e garantir as bases para uma posterior análise dos parâmetros de qualidade da água ao longo do tempo, desenvolveu-se um sistema, com a designação de AQuA, especialmente projectado para a recolha, filtragem, transformação e integração dos dados provenientes das leituras realizadas pelas estações de monitorização. Complementarmente, criaram-se neste sistema as estruturas necessárias para a realização de todas as tarefas de carácter operacional, assim como os mecanismos necessários para a manutenção e exploração de um *data warehouse* específico e consequentes estruturas para suporte ao processamento analítico de dados. O sistema AQuA é parte integrante de um projecto mais abrangente: o ODeAnA. Neste artigo, apresenta-se a contextualização do sistema, a sua estrutura principal e descrevem-se os seus principais componentes e respectivas funcionalidades.

3 - O PROJECTO ODEANA

O projecto ODeAnA pretende contribuir para um desempenho eficaz da EDIA na gestão da água na região de influência do empreendimento do Alqueva. O objectivo é reforçar as competências da empresa nesta área, dotando-a de meios tecnologicamente avançados que sustentem políticas de gestão ambientalmente correctas. A plataforma pressupõe uma abordagem multi-disciplinar, incluindo a participação de técnicos responsáveis pela gestão da água e especialistas em modelação hidrológica, qualidade da água, gestão de bases de dados e processamento analítico. Em termos básicos, o projecto apresenta competências aos seguintes níveis: monitorização, diagnóstico e elaboração de relatórios; capacidade técnico-científica em termos de análise e modelação da informação; e, definição estratégica de cenários para a avaliação de procedimentos de exploração, podendo servir de base à promoção da descrição, planeamento, tomada de decisão e gestão corrente do EFMA.

O ODeAnA integra três sistemas principais: informação, modelação e análise (Figura 1).

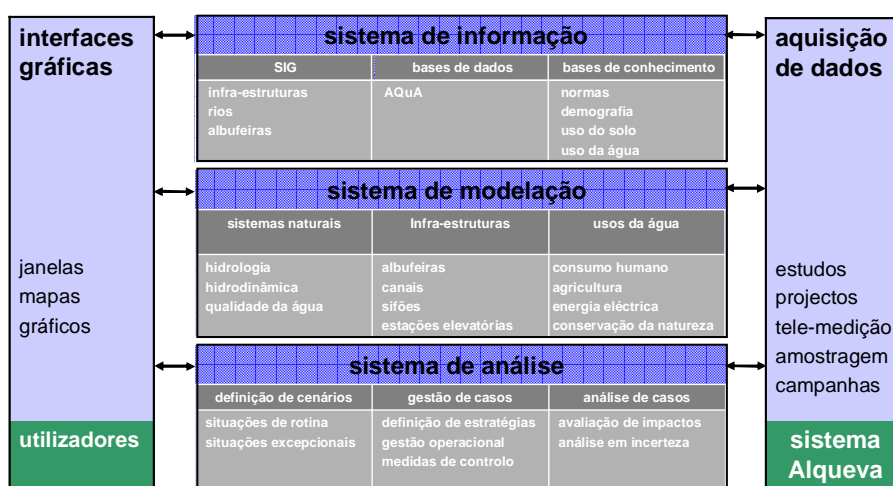


Figura 1 – Componentes do projecto ODEANA.

O sistema de informação abarca um sistema de informação geográfica, as bases de dados e os meios de difusão de informação. O sub-sistema de modelação é formado pelos modelos hidráulicos e de qualidade da água das diferentes infra-estruturas que compõem o EFMA. Os dados do sub-sistema de informação sustentam o seu funcionamento e permitem que ele forneça dados ao sub-sistema de análise, o qual está encarregue, nomeadamente, da definição estratégica de cenários e a análise de casos (Figura 1).

4 - O SISTEMA AQUA

4.1 Perspectiva Geral

O projecto AQuA tem como objectivo principal o desenvolvimento de um sistema para análise e validação de parâmetros de qualidade da água do EFMA. O sistema visa recolher os dados provenientes de leituras sobre a qualidade da água efectuadas pelas estações de recolha automática ou por técnicos de laboratório, analisando e validando esses mesmos dados e, posteriormente, integrando os dados limpos e consolidados numa base de dados projectada especificamente para o efeito. O sistema assegura mecanismos de detecção de anomalias, bem como, mantém esquemas de resolução para alguns tipos dessas ocorrências. Em termos de análise, é permitida a exploração dos dados e a geração de relatórios ou gráficos específicos. O sistema está preparado de forma a poder acolher os diversos perfis de utilização definidos – administrador de sistemas, supervisor de resultados, utilizador internos e externos e público em geral. As funcionalidades do sistema são configuradas de acordo com cada um destes perfis.

A implementação da plataforma do sistema AQuA (Figura 2) pode ser dividida em três grandes áreas: a extracção, o processamento e o armazenamento dos dados provenientes das diversas estações (AQuA Wrapper); a criação da interface Web; e, o estabelecimento das rotinas de descoberta de conhecimento consideradas mais relevantes para o caso. No que diz respeito à recolha e tratamento de dados já existentes há que referir que esta necessita de contemplar vários formatos de dados e lidar com a inconstância ao nível do número de parâmetros considerados. Mais ainda, não existe ao nível das fontes mecanismos efectivos de verificação dos dados pelo que, já foi constatado que, existem dados incongruentes face aos valores admissíveis para cada um dos parâmetros. Cabe portanto a esta parcela da plataforma, tipicamente habitada por *wrappers*, atender à diversidade e inconstância das fontes, assegurando a qualidade dos dados que passam a fazer parte do repositório do AQuA. Cada *wrapper* especializa-se numa dada fonte e lida como um conjunto específico de problemas de limpeza. À medida que são encontrados erros ou anomalias, estes são introduzidos num *log* de forma a mais tarde serem estudados, otimizando, assim, gradualmente, os processos de extracção e tratamento de dados.

4.2 A plataforma de extracção do sistema

A plataforma de extracção suporta as seguintes entradas de dados: ficheiros de texto, ficheiros em formato Excel e dados com formatos específicos provenientes de formulários (interface Web do sistema). Os ficheiros de dados recolhidos só poderão ser processados se o seu formato obedecer aos padrões estabelecidos pelo AQuA. A actual versão já processa e integra dados em formato .txt – estações – e em .xls – técnicos, sendo que estes últimos têm que obedecer a uma mesma estrutura de metadados. Complementarmente, foram já desenvolvidos alguns formulários específicos para a anotação de leituras através da interface Web do sistema, com vista a suportar no futuro a eventual transmissão de resultados a partir de plataformas remotas com acesso ao sistema.

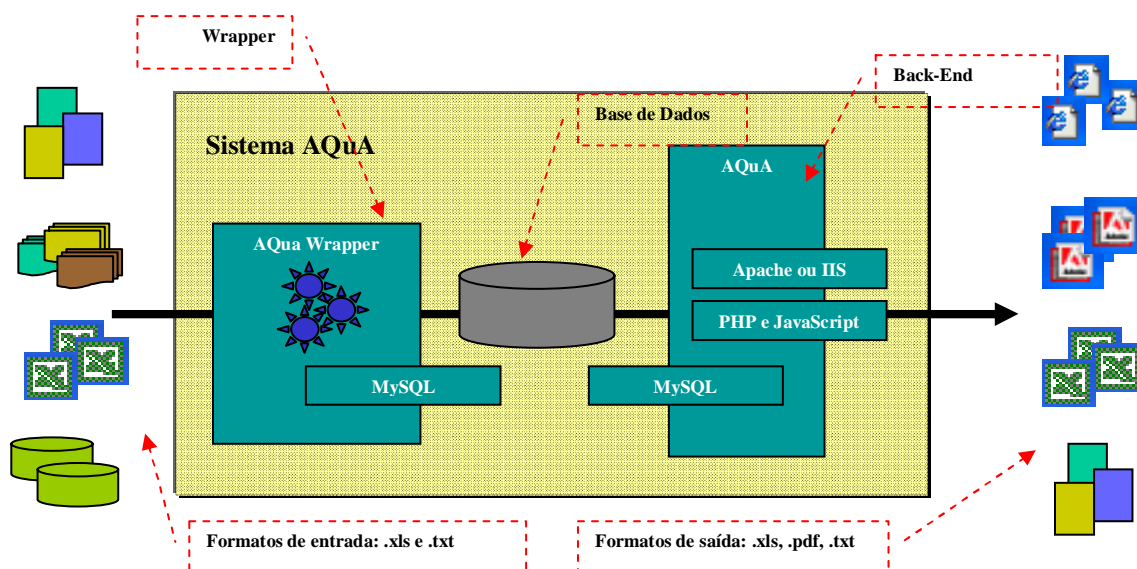


Figura 2 – Arquitectura funcional do módulo base do sistema.

Adicionalmente, é mantido um dicionário com todos os parâmetros relevantes, em termos de avaliação da qualidade da água e a sua respectiva descrição, unidade de medição, valores admissíveis (critério mínimo, critério máximo e intervalo de variação normal) e documento com explicação alargada. Para cada tipo de rede (qualidade, meteorológica e hidrométrica) existe uma matriz relativa aos parâmetros que lhe são relevantes, ao mesmo tempo que, para cada estação se efectiva o subconjunto de parâmetros que são recolhidos. Assim sendo, o trabalho de validação dos *wrappers* inicia-se com a verificação do formato e conjunto de parâmetros esperado para cada uma das fontes. Logo que esta etapa é concluída, passa-se à verificação e eventual correcção dos diversos registos. Cada valor é contrastado com o tipo de dados e a gama de valores estabelecida para o parâmetro em questão. Caso seja detectada uma anomalia e o *wrapper* disponha de conhecimento para resolver a mesma, o esquema de resolução é aplicado de imediato. A situação é registada ao nível do *log* de erros, etiquetando o problema como solucionado, e os dados prosseguem o processo de migração para a base de dados. Caso contrário, a situação é registada ao nível do *log* de erros, etiquetando o problema como não solucionado, sendo os dados excluídos. Todavia, os dados recusados são armazenados numa zona particular da base de dados, aguardando uma eventual solução por parte dos administradores do sistema.

A recolha dos ficheiros contendo os dados relativos às leituras sobre a qualidade da água numa determinada zona do sistema Alqueva é realizada via FTP a partir da máquina do sistema que recebe os dados directamente das estações. Os diversos ficheiros são copiados das respectivas fontes para efectuar o respectivo processamento ao nível do AQua, sendo colocados numa zona privada do sistema designada por `\aqualmontante\<rede>`, em que `<rede>` representa a rede na qual foram efectuadas as leituras. Os dados são processados pelo módulo de *wrapping* do AQua que os integra na base de dados do sistema de acordo com os requisitos de qualidade exigidos. No fim do processamento os ficheiros processados são transferidos para uma segunda zona privada do sistema designada por `\aqualjusante\<rede>`. Somente no final de todo este processo se realiza a efectiva remoção ao nível da fonte garantindo-se sempre que este processo não conduziu a uma situação na qual se verificou algum tipo de perda de informação. Nesse sentido, foram criados mecanismos específicos para fazer a recuperação de dados em situações de falha.

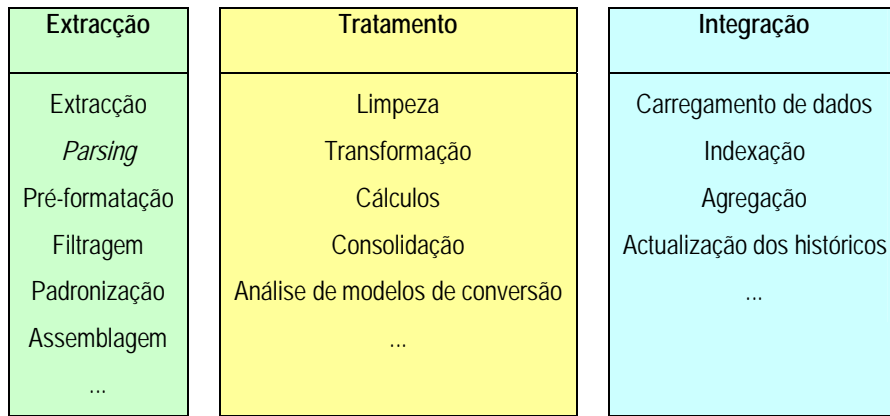


Figura 3 – Tarefas principais do sistema de *wrapping*.

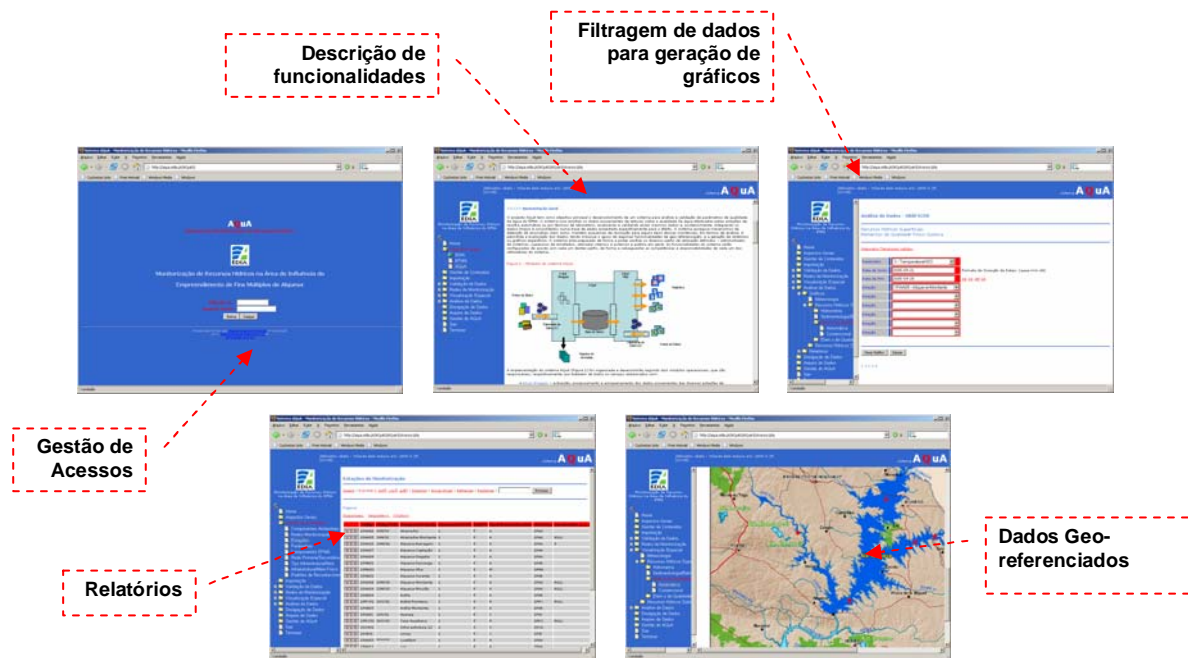


Figura 4 – Exemplos de écrans do sistema de interface do sistema.

4.3 A Interface do Sistema

A interface Web (Figura 4) foi modelada tendo por base as necessidades operacionais e analíticas básicas do sistema. Esta plataforma assegura a introdução, consulta e análise dos dados relativos às redes de qualidade, meteorológica e hidrométrica. Os diversos parâmetros de referência são tidos em conta, mantendo-se uma breve descrição acerca dos mesmos, bem como, acerca dos valores admissíveis em cada caso, sendo os novos dados validados em concordância. Assegura-se ainda a análise destes mesmos dados recorrendo a um sistema de informação geográfica, elaborando gráficos relativos à evolução dos distintos parâmetros e permitindo a exportação de dados e resultados, bem como, a elaboração de pequenos relatórios. A plataforma assegura também um sistema de autenticação dos utilizadores, que acompanha a definição dos diversos perfis de utilização estabelecidos, e de registo das actividades realizadas dentro da plataforma ao longo do tempo.

4.4 A Componente Analítica do Sistema

A componente analítica do AQuA tem como objectivo principal abarcar todos os processos de análise convencional, através da geração de estatísticas, relatórios e gráficos, bem como dos processos de análise ditos não convencionais suportados pelas estruturas multidimensionais de dados asseguradas pelo *data warehouse* do sistema (Figura 5). Com vista a atingir tais objectivos, esta componente mantém um conjunto de *data marts*, cada um dos quais relativos a um tipo de rede (qualidade, meteorológica e hidrométrica), envolvendo toda a informação que se encontra disponível e que é considerada relevante em termos de processamento analítico.

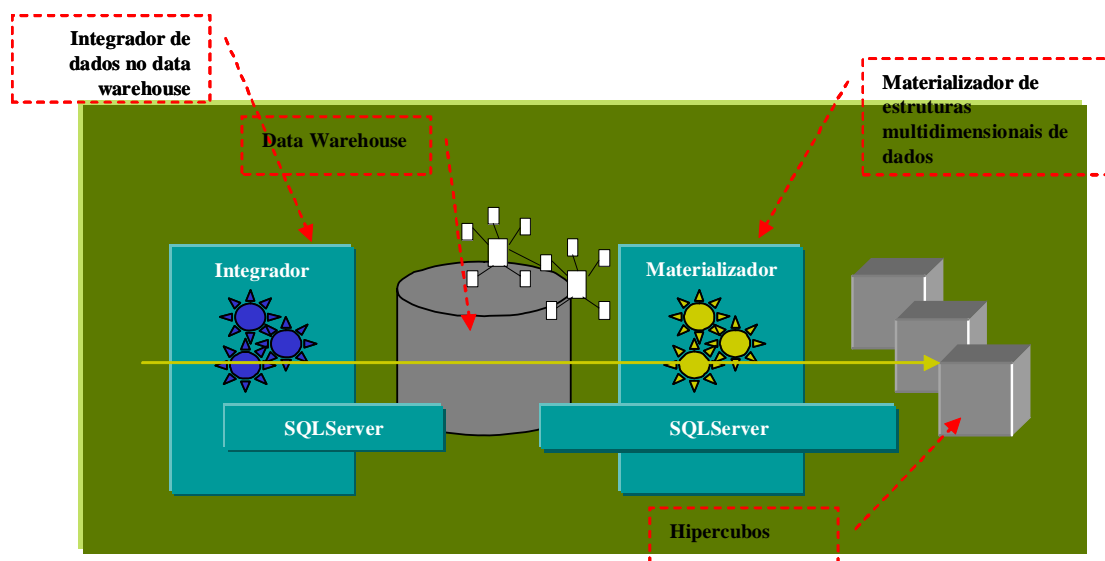


Figura 5 – Arquitectura funcional do módulo analítico do sistema.

A Figura 6 apresenta de forma simplificada o esquema relativo ao *data mart* sobre a qualidade da água. Tal como se pode observar, as perspectivas de análise dizem respeito à origem dos dados em termos de sistema e rede a que pertence, estação que efectuou a monitorização, ficheiro de onde foi extraído o registo e seu tipo e ainda, à dimensão temporal do sistema.

O materializador de estruturas multidimensionais de dados baseia-se nestes *data marts* para efectuar a construção dos hiper cubos. Mais tarde, estes hiper cubos servirão de base à aplicação de processos de descoberta de conhecimento. Algumas destas dimensões são partilhadas com os *data marts* relativos às outras redes, apesar de cada uma delas poder integrar dimensões que lhe são únicas. As medidas mantidas são as consideradas relevantes para avaliação da rede em questão. No caso concreto da rede de qualidade da água, medição automática, interessa manter o registo dos valores da condutividade, oxigénio dissolvido, pH, temperatura, turvação e potencial Redox. Em termos de esquema final do *data warehouse*, dada a partilha de dimensões, os vários *data marts* acabam por compor uma constelação.

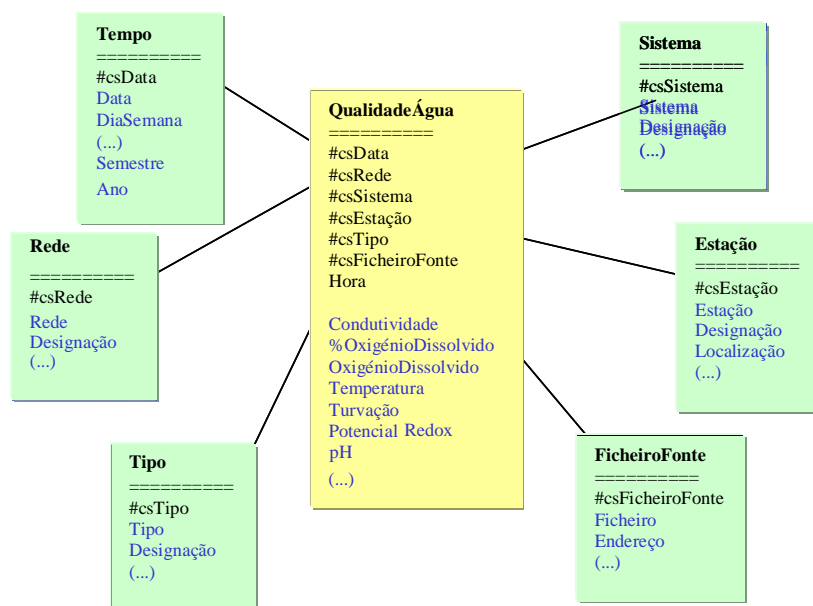


Figura 6 – Esquema do *data mart* de qualidade da água.

O *data warehouse* do sistema, assim como os hiper-cubos projectados, estão já preparados de forma a assegurar as futuras actividades de *data mining* já planeadas. Nesse processo, existirão três linhas de orientação primordiais:

- A mineração de associações entre os distintos parâmetros de qualidade da água e a construção de modelos de previsão, tentando-se antever eventuais situações de risco.
- Os ficheiros de *log* relativos ao acesso ao sítio Web servirão para perscrutar o trabalho realizado pelos utilizadores, impulsionando a criação de um outro *data mart* relativo aos acessos à interface Web (pode-se analisar os perfis dos utilizadores e a evolução das necessidades de recolha de índices ao nível das estações).
- Os ficheiros de *log* relativos aos erros dos *wrappers* que serão minerados de forma a agrupar, interpretar e classificar novos erros e a encontrar potenciais esquemas de resolução, actualizando as bases de conhecimento dos respectivos extractores.

5 - CONCLUSÕES E TRABALHO FUTURO

O protótipo actual cobre os requisitos e funcionalidades estabelecidos para a fase de arranque do sistema, em particular ao nível da integração e exploração de dados. Durante o desenvolvimento do sistema teve-se sempre em consideração uma perspectiva de aplicação genérica, tanto ao nível das plataformas de utilização como ao nível das funcionalidades de análise implementadas. Nesta fase considerou-se já importante a implementação de alguns mecanismos de autenticação e de controlo de acessos com vista a assegurar a segurança e a privacidade dos dados. Para demonstrar a instalação e utilização “universal” do actual sistema protótipo desenvolveu-se um sítio provisório na Internet, de forma a poder verificar-se o seu desempenho em termos de acesso remoto e de aplicação aos requisitos operacionais de um ambiente real.

Após esta primeira fase de desenvolvimento do protótipo partir-se-á para a implementação efectiva do sistema, cobrindo funcionalidades ainda não abrangidas pelo protótipo desenvolvido, optimizando os

sistemas de análise e efectuando testes de *stress* e robustez em tempo real. Futuramente, pretende-se alargar o espectro de aplicação do AQuA para que se possa realizar:

- O acolhimento de dados provenientes das restantes redes do sistema, nomeadamente a meteorológica e a hidrométrica.
- A reavaliação das vistas de utilização e a completa caracterização de todas as redes integradas no sistema, efectuando o levantamento de todas as potenciais consultas de dados e futuros perfis de utilização.
- A definição e a implementação de um sistema de protocolo para a comunicação de dados provenientes de recolhas efectuadas por estações automáticas ou por técnicos de laboratório.
- O projecto e desenvolvimento de mecanismos de autenticação e segurança do sistema mais efectivos e abrangentes.
- A inclusão de um sistema de *clickstream* que permita o estudo dos padrões de navegação no sítio do AQuA.
- O projecto e a implementação de um módulo para descoberta de conhecimento.

6 - AGRADECIMENTO

Os autores agradecem à EDIA o apoio financeiro concedido para a realização deste estudo.

7 – REFERÊNCIAS BIBLIOGRÁFICAS

- CALVANESE, D., GIACOMO, G., LENZERINI, M., NARDI, D. E RICCARDO ROSATI, R. (2001) *Data integration in data warehousing*, International Journal of Cooperative Information Systems, 10(3):237-271.
- DQA (2002) Directiva 2000/60/CE do Parlamento Europeu e do Conselho, Jornal Oficial das Comunidades Europeias, 2000.
- HERNANDEZ, M. A. E STOLFO, S. J. (1998) *Real-world data is dirty: Data Cleansing and the Merge/Purge problem*, Journal of Data Mining and Knowledge Discovery, 2(1):9-37.
- KIMBALL, R. (1996) *Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*, John Wiley & Sons.
- NIEMI, T., NUMMENMAA, J. AND THANISH, P. (2001) "Constructing OLAP Cubes Based on Queries", in J. Hammer, editor, DOLAP 2001, ACM Fourth International Workshop on Data Warehousing and OLAP, ACM, 9-11.
- RAHM, E. E DO, H. H. (2000) *Data Cleaning: Problems and Current Approaches*, Bulletin of the Technical Committee on Data Engineering, vol. 23, n.4, pp. 3-13, IEEE Computer Society.
- WIEDERHOLD, G. (1992) *Mediators in the architecture of future information systems*, in IEEE Computer 25, 3, pp. 38-49.
- WINKLER, W. E. (2004) *Methods for evaluating and creating data quality*, Information Systems, In Press.
- WOOLDRIDGE, M., WEISS, G. E. CIANCARINI, P. (Ed.s) (2002) *Agent-Oriented Software Engineering II*, Springer Verlag Lecture Notes in Computer Science Volume 2222, February.