Mapping Water Quality Information for Public Presentation: GIS at Tucson Water

Jesse Gordon, Senior Database Developer, TPMC, Scituate MA

Marie Pearthree, Assistant Director, Tucson Water Department, Tucson AZ

Michelle Frey, Vice President, McGuire Environmental Consultants, Denver CO

David Miller, Manager of Software Engineering, TPMC, Scituate MA

1. ABSTRACT

The City of Tucson Water Department presents real-time water quality information on their web site for use by the public and the press. City residents use the map-based interface to zoom in to their neighborhood, view the nearest distribution sampling point and details of its water quality, or view a summary of the water quality in their neighborhood or the whole city. The water quality information is updated daily, and results are posted for a dozen analytes the day after the data is collected.

The intent of the project was to develop an interface that would allow utility personnel and consumers in the Tucson area to access data of high quality in near-real time. The initial requirements of the system included the ability to quality-assure data in an automated fashion using statistical methods, and a map-based interface that could access a database and perform aggregations and analysis upon request. The system was implemented using a customized ESRI ArcIMS interface for the map-based components. All of the data presentation is accomplished via Java servlets and applets on an IBM WebSphere Application Server accessing data from an Oracle database server.

The system was made public in May 2001. It was warmly received by the utility personnel, the media and the public. The web site has received several hundred visits per week. The web site allows users to answer the following questions based on the most current data available:

- ***What is my water quality?*** A detailed map down to the property parcel level lets residents find their street and the nearest water sources.
- ***How does my water compare to the rest of the city and the country?*** Each neighborhood is compared to all others on a color-coded map, and national water quality standards are detailed.

- ***Where does my water come from?*** A static depiction determined by hydraulic models illustrates the complex process of mixing ground water from wells located throughout the city with recharged water from the Colorado River aqueduct.

The system is designed to allow users to specify which analyte they wish to investigate from a list established by the utility. The utility specifies the period of record, how the data should be aggregated and also the data is presented (tabular, graphical or maps).

This paper will discuss the challenges encountered, solutions applied and applicability to other drinking water utilities.

2. BACKGROUND

The City of Tucson was the largest U.S. city to depend entirely on groundwater, until May 2001, when the city began "blending in" water recharged from the Colorado River aqueduct. Tucson is one of the fastest-growing metropolitan areas in the U.S., with a population that has grown by 26% in the last 10 years to 850,000, with a projection to reach one million by 2010 [US census figures]. Tucson Water services about 640,000 customers for both drinking water and wastewater.

The fast-growing population drawing water from Tucson's underground aquifer means that the aquifer is being depleted, causing a measured subsidence in surface levels of several inches per decade – Tucson is slowly sinking. The subsidence is particularly severe in the center of the city, where many of the groundwater wells are concentrated. The urgency to replace those wells, and to avoid sinkholes from further subsidence, was the impetus to initiate the aqueduct project. Simultaneously with the aqueduct construction, which was completed in the early 1990s, new well fields were drilled outside of the center of the city, partially alleviating the problem. The most important event at the May 2001 switchover was that the central city wells were NOT turned on for the summer season, as they had been in past years.

The blending-in of recharged Colorado River water will be phased in over a period of several years, to allow both the public and the distribution systems to adjust to any potential changes in the character of the water. The blending of water occurs in two ways. First, the river water is poured into leaching fields, and over a period of about two years, filters into an underground aquifer – called "recharging." That aquifer is then drawn upon and the resulting river water is further mixed with local aquifer water in the distribution system to make it closer to local pH and mineral content – called "blending." The overall process is known as the "Clearwater" project.

As part of an EMPACT grant from the USEPA, water quality information is provided to the public from a number of different sources. These include:

- Tucson's water quality lab presents summaries of its LIMS database on the Water Department web site, comparing water quality information before and after the blending began at 400 distribution sampling points and points-of-entry.
- Tucson collects "grab samples" in the field from several dozen collection points per day, and the results are presented on the web site the day after the tests are completed.
- Tucson Water's web site presents summaries of hourly water quality readings from 22 automated sensors, installed under an EPA EMPACT grant on water pipes around the city.

The data is combined and presented from each of the sources and is presented on a public web site. This system has helped to establish a partnership between the utility and its customers.

3. WEB SITE

The primary goal of the map-based section of the web site is to answer the questions, "What is my water quality?" (in the viewer's neighborhood) and "What is Tucson's water quality?" (for city-wide comparisons). The map-based sections of the web site are split into two sections according to those two questions, http://www.ci.tucson.az.us/mw/ ("mw" for "my water") and http://www.ci.tucson.az.us/cw/ ("cw" for "city water").

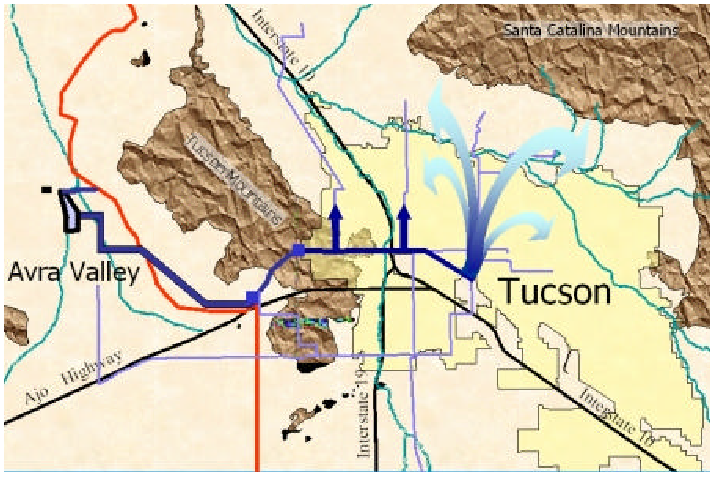


Figure A

A map-based method was considered to answer a third question, "Where does my water come from?" but the modeling required is too complex to be provided in a dynamic manner. Specifically, the source of each area's water varies depending on the time of day, the season of the year (as wells are turned on and off), and past and current water usage. Had this question been answered graphically, percentage terms would have been used – for example, 20% came from recharged Colorado River water, 30% came from well A, and 50% came from well B. We concluded that calculating accurate percentages for each area would be difficult and would not give most viewers any more accuracy. A more general description of how the distribution system works and how the multiple sources of water are blended was determined to be more practical. This information is provided statically based upon hydraulic models (Figure A), at http://www.ci.tucson.az.us/water/water_resources/clearwater/am_i_receiving/am_i_receiving.htm This section of the web site is static text and graphics and won't be discussed further here.

3.1. City Water – Map Navigation

In the "city water" section, the viewer is initially greeted with a schematic map of the City of Tucson and its surrounding areas to the limits serviced by Tucson Water. The city is divided into 18 "Water Quality Zones," which represent different pressure zones in the distribution system within the main city system, and eight isolated system zones outside of the main system. The zones are indicated by the numbers with the yellow glow on the map, and are referred to as "Zone 1" through "Zone 18." Note that the schematic map has gaps between the zones – those represent areas within the city where residents are serviced by companies other than Tucson Water or by private wells. The water quality zones – the varying pressure zones – were chosen as the primary metaphor for representing the city's water system as they are reasonable boundaries within the system.

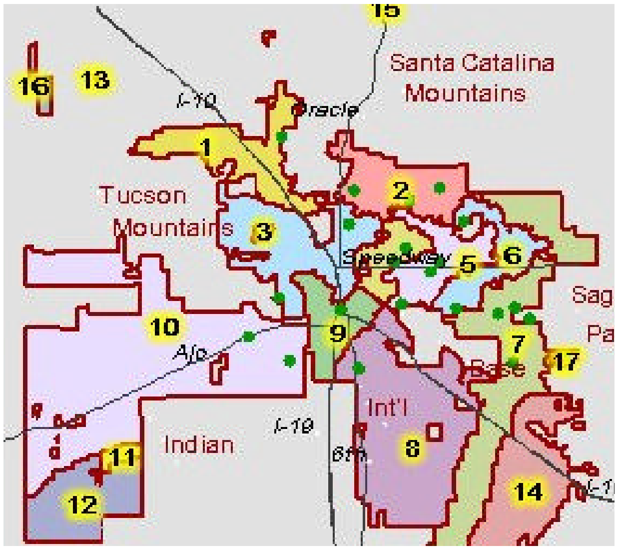


Figure B

Figure B represents a fully functional dynamic map – the user can pan (shift from east to west and north to south, or any direction in between) and zoom (magnify inward or change to a larger scale outward) using mouse-based icons. The map will show meaningful data at any scale from several dozen miles to just a few yards. Additional data layers are shown as the viewer zooms in – for example, major streets are shown at a scale of 2 to 5 miles across the screen (Figure C), and individual properties are shown at a scale of about ¼ mile across the screen (Figure L below). The scale is shown on the bottom of each map.

To make navigation readily apparent to viewers, Tucson's major highways and Interstates appear on the initial screen, as well as large landmark names like the Santa Catalina Mountains. Each water quality zone is color-coded differently, and those colors are consistently shown as viewers zoom in to their neighborhoods. Hence viewers can readily determine the water quality zone in which they reside, and can request water quality information for their own zone, and compare to other zones.

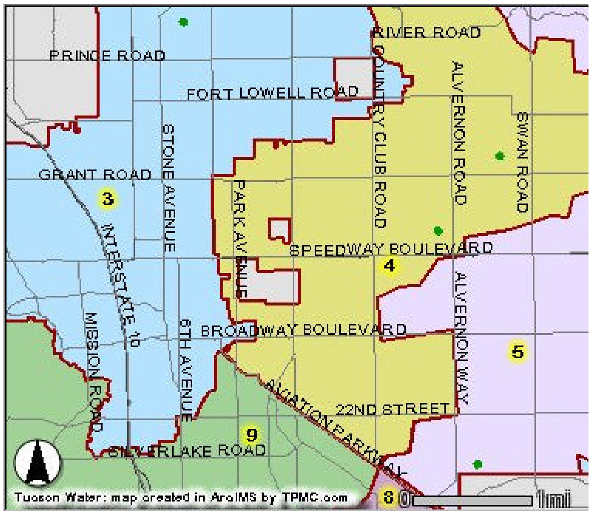


Figure C

The green dots represent 22 automated sensors where hourly water quality readings are taken and reported to a central database. These points are referred to as "EMPACT points" because the sensors were installed under an EPA EMPACT grant. They are referred to as EP1 through EP22 in the data tables available on the web site. Only two of the EMPACT sensors are currently operable (as of January 2002); the rest will come on-line during 2002. In the interim, "grab samples" are made manually at the other 20 EMPACT points, so that data is currently available for all 22 points.

The dynamic map has four modes:

- ***Pan:*** shift location without changing scale, by dragging with the mouse;
- ***Zoom in:*** show more detail, either by clicking the mouse, which zooms in by a factor of two and re-centers on the point clicked, or by dragging the mouse to outline a specific area to display at a larger scale;

- ***Zoom out:*** show more area of the city, by clicking the mouse while the zoom-out mode is set, which zooms out by a factor of two, or by clicking "Reset," which returns to the initial screen in Figure B;
- ***Identify:*** show water quality information in one water quality zone or at one sampling point.

When using Identify mode, we pre-determine what information will be shown, based on which map is currently displayed, and at what scale. In the case of the initial map in Figure B, "Identify" always displays information about the EMPACT points at all scales. If the viewer clicks near one of the 22 green EMPACT points, a pop-up window with a line graph appears, as in Figure D. If the viewer clicks more than 10 pixels from any EMPACT point, they're notified that on this map they must click near EMPACT points to identify them.

3.2. City Water – Graphs

Figure D indicates the automated temperature readings at one EMPACT point. The actual hourly readings are summarized into one daily average by a daily batch process (discussed below with Figure R). The date axis shows the most recent 60 days of data, pushing old data off to the left as new data becomes available. Below the line graph on the pop-up screen appears the same data in tabular form, showing each date and the temperature value for that date. The viewer has no access to data older than 60 days (although that data is stored on the database for other purposes).

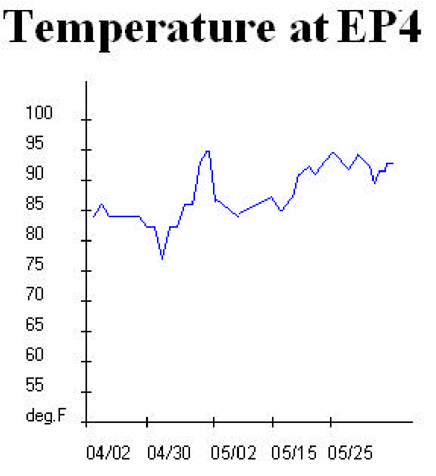


Figure D

Water Quality measures: EMPACT Point EP4

Chlorine, mg/L	0.94	MAY 08 2001
Fluoride, mg/L F	0.52	MAY 01 2001
Nitrate, mg/L N	2.2	MAY 01 2001
Sodium, mg/L Na	42.0	APR 11 2001
TDS, mg/L	411.0	MAY 08 2001
Temperature, deg.F	95.0	MAY 08 2001
pH, Std. Units	7.7	MAY 08 2001

Figure E

Below the line graph on the pop-up window are links to the line graphs for other EMPACT analytes for the currently selected EMPACT point: chlorine residual, pH (acidity), and TDS (total dissolved solids). Those four analytes were chosen for line graphs because they are measured by the automated EMPACT sensors – i.e., all the measured analytes are shown in this case (which is not the case elsewhere). We select which line graph to show based on what the viewer has most recently requested – hence if the viewer clicks on "chlorine" and then returns to the map and clicks on a different EMPACT point, its chlorine line graph will be shown first. The intent was to provide a means of surveying one analyte across the city, as well as a means of surveying all analytes for one EMPACT point.

An additional link for "Most recent readings" is presented with each line graph Clicking there pops up a window showing a table like the one shown in Figure E. Here, we present all of the grab-sample analytes as well as the automated sensor analytes for one EMPACT point. We present just one figure for each analyte, representing the single most recent available reading. Hence for sensors which are operating in automated mode, the four automated analytes will always show yesterday's data (since they are recorded every day and batch-processed overnight) and will show older dates for analytes which are only collected by manual grab sampling. The grab-sampling schedule for EMPACT points is arranged to get at least one reading per analyte per month at each sampling point.

Tucson Water collects grab samples for numerous analytes which are not shown on Figure E. The number of analytes was limited to those with the most public interest, and for which data is regularly collected. Tucson Water could add new analytes to the list shown in Figure E, if desired, which would immediately result in a longer list

On the list of the most recent readings, clicking on one of the four analytes reported by automated sensors returns the viewer to a line graph (Figure D). Clicking on the other analytes brings up a bar graph (Figure F) with one summary value per month. We show monthly summaries for these analytes instead of the daily summaries as in Figure D because these analytes are collected by manual grab sampling and therefore daily data would not be meaningful. The grab sampling is done on an approximate one-month rotation, so each bar represents one or sometimes two readings, and in some cases a bar is absent for a month.

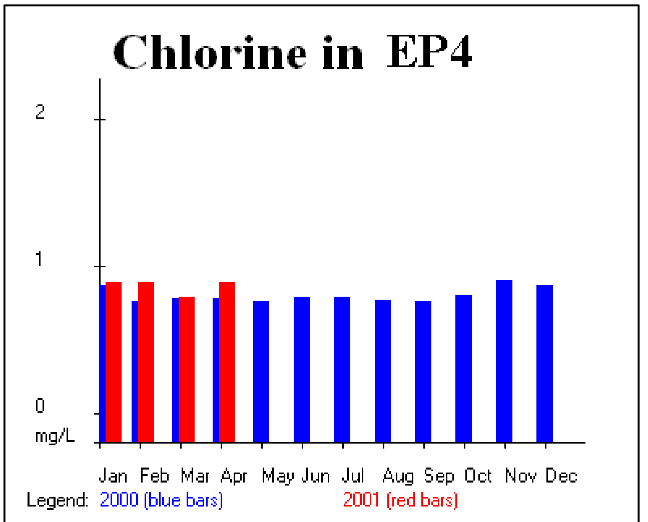


Figure F

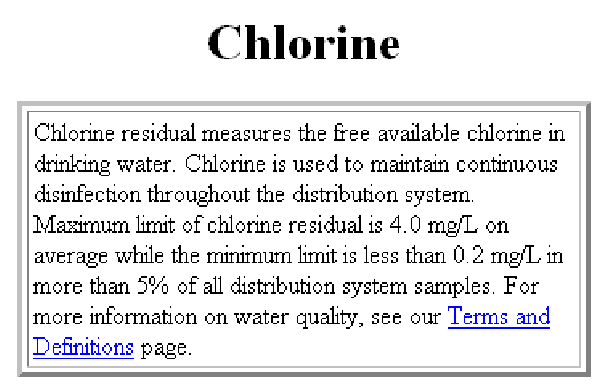
The bar graph color-codes the current year in red and the previous year in blue, to allow direct year-by-year comparison. Consistent data was not available prior to 2000, so we limited the data to January 2000 to the present. We recently added a third set of bars for 2002, and from January 2002 onward, three year comparisons will be available. Note that the date is fixed – it is always January to December. Hence the amount of data shown varies from 25 to 36 months, depending on the month of the year (for the bar graph in Figure F, prior to 2002, the amount of data varied from 13 to 24 months).

Note too that the vertical axis is similarly fixed – for chlorine, the axis always shows a scale from 0 mg/L to 2 mg/L, regardless of the values of the data for the particular sampling point being shown. This allows viewers to recognize visually when an analyte had a high reading compared to other sampling points – i.e., a tall bar on one sampling point means the same thing as a tall bar on any other sampling point. This differs from the typical method of showing bar graphs, where the vertical axis is adjusted depending on the values of the data – we rejected that because it would require the viewer to read the numbers rather than to make a visual comparison.

We spent considerable effort in coming up with reasonable vertical axis limits that would cover foreseeable analyte values. A shortcoming of pre-determined axis values is that extremely high values will appear "off the chart." We address that problem by showing the actual analyte readings in tabular numeric form below each bar graph. Another shortcoming is that low values will not show bars at all, and in the case of Figure F, zero values would not be differentiable from a lack of a reading. We address that problem by making the vertical axis extend slightly below the "lowest value" of the axis, so a zero reading shows a short bar, while a lack of any reading shows a blank for that month. To determine the axis limits, we used the following criteria:

- Show tall bars when the values were actually higher than average, i.e., choose the center point and high point so that typical values fall around the center of the bar height;
- Include the zero point as often as possible, but not if it makes the bars too tall (for example, in Figure D, a zero degree Fahrenheit temperature is not applicable to Tucson Arizona);
- When possible within those criteria, show the regulatory limits for the analyte.

The regulatory limits are defined for the viewer on pop-up "blurbs" (Figure G). These blurbs are available from all tabular data, which appear under the line graphs and bar charts, and are linked from other locations as well. The intent is to provide background about the analyte, define why it appears in the water supply, and to establish the regulatory limits for comparison to the sampled values.



Chlorine

Chlorine residual measures the free available chlorine in drinking water. Chlorine is used to maintain continuous disinfection throughout the distribution system. Maximum limit of chlorine residual is 4.0 mg/L on average while the minimum limit is less than 0.2 mg/L in more than 5% of all distribution system samples. For more information on water quality, see our Terms and Definitions page.

Figure G

3.3. City Water – Colorized Maps

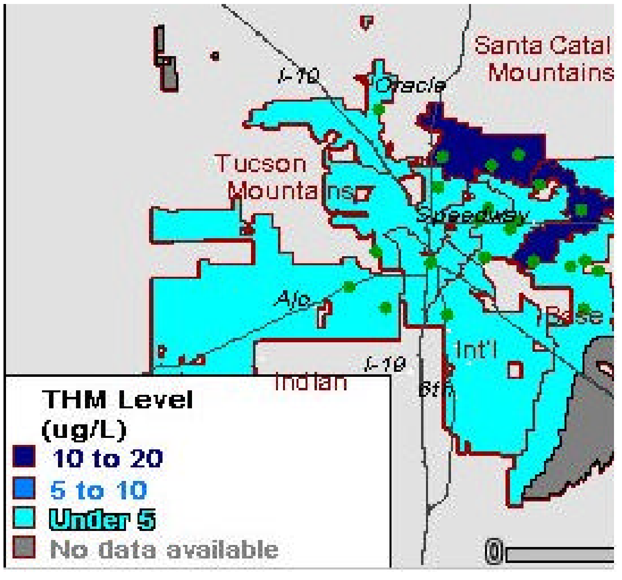


Figure H

The above methods allow the viewer to "drill down" to see water quality information across the city in detail. We also provide a means for direct comparison between water quality zones. There are a series of buttons, one for each selected analyte, which bring up "colorization maps" such as Figure H. On these maps, we re-color the water quality zones with one of four colors:

- Dark blue for a "high" level for the analyte in the specified zone;
- Medium blue for a "medium" level;
- Light blue for a "low" level;
- Gray if no samples were taken.

As with the bar charts and line graphs, we pre-determine which values are indicated as high, medium, and low. The alternative here would have been to colorize, say, the 6 zones with the highest readings as dark blue, and the 6 lowest zones as light blue, and then report the cutoff values accordingly. We rejected that method because it would always imply that some zones were "too high," and we wanted to avoid an appearance of being out of compliance with regulated water quality levels. The benefit of pre-determined cutoffs is that they do not change over time – a viewer can come back after a period of absence and the colors will always mean the same thing.

We present colorization maps for the following analytes:

- Coliform (total, not E. Coli)
- Chlorine (residual measured as mg/L)
- TDS (total dissolved solids)
- Hardness (measured as mg/L CaCO3)
- pH (acidity)
- Nitrate (measured as mg/L N)
- Sodium (measured as mg/L Na)
- THM (trihalomethanes)
- Fluoride (measured as mg/L F)

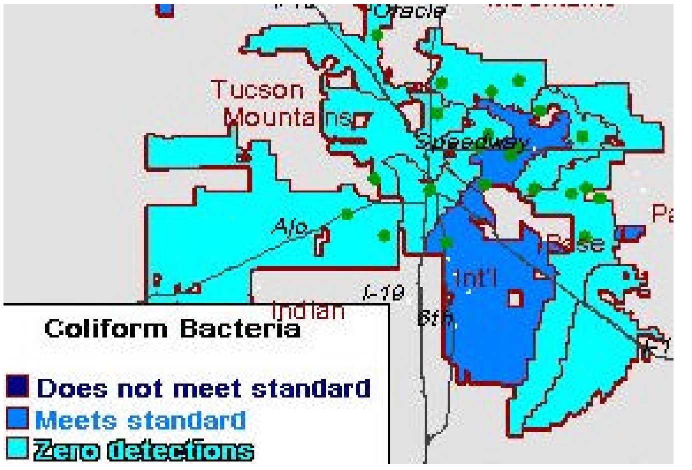


Figure I

As with our previous analyte pre-selections, we chose this analyte list based on public interest and availability of data. The database allows the presentation of many other analytes using the same method, but in this case Tucson Water must re-program the web page to add new analytes, instead of simply adding them to a list. This is necessary because the web page has graphic buttons to link to the colorization maps, and the data summary for colorization is output to HTML objects. The values for each water quality zone are determined by averaging all analyte readings for the previous 12 months.

The colorization map for Coliform Bacteria (Figure I) presented several issues due to public sensitivity to positive coliform readings. All the other colorization maps simply report numeric values, whereas for Coliform, we report presence or absence of any coliform at all. The regulations state that any water quality zone which has two positive readings in any one-month period is out of compliance, and any water quality zone which has one positive reading in any two subsequent months is out of compliance. Those regulations apply to isolated zones; the main system is allowed a 5% positive rate. Hence there are three possible states, as shown on Figure I: out of compliance (dark blue), in compliance but not zero (medium blue), and zero detections (light blue).

The map shown in Figure I is not shown on the Tucson Water web site. We decided to eliminate the “zero detection” reading and limit the map to two colors: dark blue for “out of compliance” and medium blue for “meets standard.” Showing zero detections would beg the question about why some water quality zones did NOT have zero detections while others did. Since water quality zones which are out of compliance must shut down the distribution system in that zone until the coliform presence is eliminated, there is never a case where zones will report as out of compliance except for one day – hence this map always shows a uniform medium-blue across the entire city.

When showing the colorized maps, the “Identify” function changes meaning. Clicking on the colorized maps pops up a window showing the monthly bar chart of values for the selected analyte averaged across the water quality zone. The bar chart is similar to Figure F, and is discussed further with Figure N below. The 22 green EMPACT points are not eligible for identification when showing the colorized maps (viewers click on “Reset” to return to the uncolorized map). We removed them from later versions because people were confused by their presence.

Note that the colorized maps are still fully functional dynamic maps, i.e., viewers can zoom and pan as with the uncolorized maps. The water quality zone colors are retained as viewers zoom in, except that streets and property parcels overlay the background zone color at higher magnification.

The colorization maps were the most complex technical task of the Tucson Water web site. The technical means of accomplishing them requires:

- An invisible data layer where each water quality zone has an indeterminate color; this data layer is made visible when an analyte is selected for colorization, after the zone colors are determined;
- An initially visible layer of the water quality zones using the standard colors; this layer obscures the colorization layer so that it can be shown in one clear transition and so that the uncolorized layer can be turned on and off;
- A hidden HTML frame which reads an HTML file containing the zone values for the analyte; this HTML file is rebuilt daily from the previous 365 days of readings (hence the colors could change daily, if the analyte values fluctuate enough).;
- A JavaScript delay, to allow time for the HTML file to fill in the hidden HTML frame. This is necessary because there is a non-trivial amount of time for reading in the file, and without the delay, the colorization drawing would proceed before the HTML file was read;
- The legend (numeric levels for each color) is drawn piecemeal as shape and text objects in the “acetate layer,” a readily-programmable image that lays on top of the map layers. This allows Tucson Water to change the cutoff levels by editing a “soft-coded” file instead of, say, a hard-coded image.

3.4. My Water – Scale-Dependent Maps

The maps intended to answer the question, “What is my water quality?” are scale-dependent: when showing a broad area, they link to charts by water quality zone, and when showing a smaller detailed area, they link to charts by sampling point.

Like the City Water map, the initial map shows all of Tucson (Figure J). At this scale, clicking to identify a point will display data summarizing the entire water quality zone (Figure M below); individual sampling points are not shown (as they are in the analogous Figure B). To differentiate between Figure J and Figure B, we changed the border colors between zones, to provide a visual cue. Other than that, the initial maps are identical.

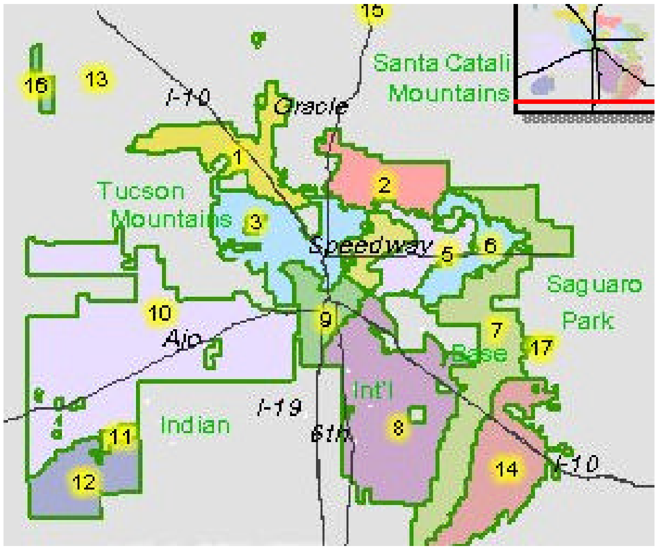


Figure J

Zooming in from the initial map results in Figure K, which has a scale about four times more detailed than the initial map. At this scale, we show water quality sampling locations: red for distribution system sampling points and blue for points-of-entry sampling points. Tucson Water samples from 258 wellhead sampling points (called "SPs") and from 160 points-of-entry (called "POEs").

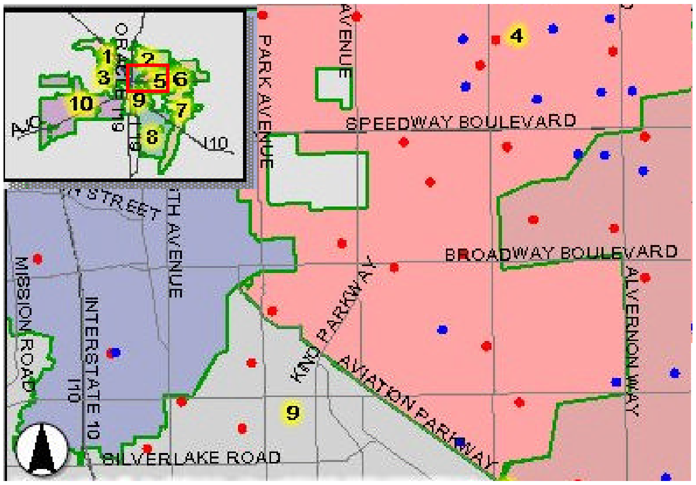


Figure K

Zooming in further shows additionally detailed data layers, down to individual property parcels (Figure L). At the most detailed level, the property parcels are shown in olive green while the streets show the background color of the water quality zone, which is the same color as on the initial map. Hence there is a visual cue at all scales about which water quality zone is being shown. In addition, the water quality zone number is shown as a yellow-glowing numeral, visible in the center of Figure L.

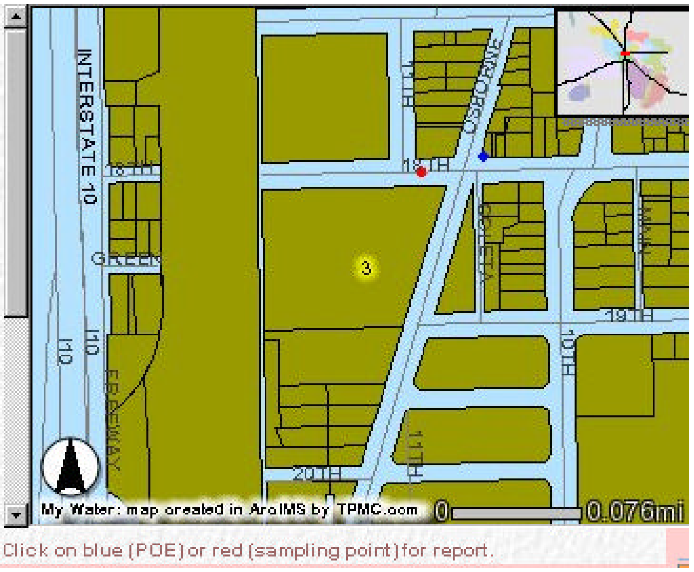


Figure L

At more detailed layers, the red SPs and blue POEs are shown in a position accurately reflecting their physical location. Interested residents could locate the actual sampling locations in their neighborhood by noting the exact location on the most detailed map – a resident physically inspecting that location would find a manhole cover or standpipe.

At all scales where the red SPs and blue POEs are shown, clicking to identify a point will show water quality data for that particular sampling point. The intent is that residents can find the sampling points nearest to their homes and interpret the water quality readings at those points as the water quality readings of their own tapwater. That interpretation is not strictly true, because water mixes in the distribution system, and water flow is one-directional, so if a home is upstream from a wellhead, it receives no water from that well even if it is only 10 yards away. Nevertheless, that interpretation is a reasonable estimate of water quality in one's neighborhood.

Note that a "locator map" is shown in the upper left of Figure K. That map indicates the area of the city that is displayed on the main map, with the red box indicating the perimeter. At more detailed scales, a red point on the locator map indicates the relation to the whole city – visible in the locator map at the upper right of Figure L.

3.5. My Water – Data Charts

Clicking on a water quality zone from Figure J pops up a tabular chart, Figure M. This table indicates the average value for each analyte for the last 12 months, the "lower limit" and "upper limit" representing the range of values, and the number of samples that make up the average.

Zone 5 Water Quality Summary

The information in this table is based on water quality monitoring results for the most recent 12-month period.

Water Quality Measure	Water Quality Measure Results			
	Average	Lower Limit	Upper Limit	Number of Samples
Chlorine, mg/L	0.83	0.48	1.14	483
Coliform Bacteria, Percent Positive	0.22	N/A	N/A	464
Fluoride, mg/L F	0.41	0.15	0.65	105
Hardness, mg/L CaCO3	103	73	125	379

Figure M

The lower limit and upper limits are statistical figures – they are not the single lowest and single highest reading for the analyte. Showing those readings emphasized statistical anomalies and instrument errors, so we decided instead to use a percentile range. The "lower limit" figure is actually the value at the 5th percentile of all readings, and the "upper limit" figure is the 95th percentile of all readings. We also tried out the 10th percentile and 90th percentile, but found those to be less informative of the range of values. Note that coliform bacteria has no meaningful percentile limits, since it measures only the percentage of positive readings.

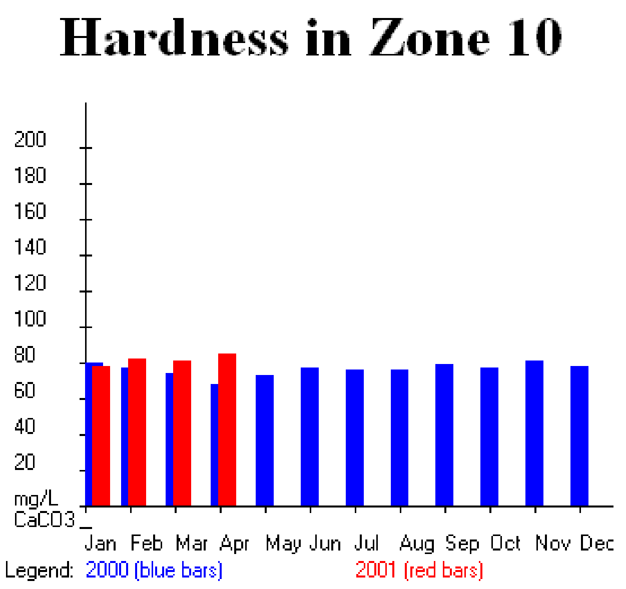


Figure N

Clicking on an analyte name in the table in Figure M brings up a bar chart, Figure N. This bar chart is analogous to the bar chart in Figure F, except that here it represents a summary of all values in one water quality zone instead of from one EMPACT point. The fixed scale and 25-to-36 month time axis are analogous in both bar charts. Note that the vertical axis uses the same fixed scale on the bar charts for both individual sampling points and zone summaries, allowing for direct visual comparison.

At more detailed scales, such as in Figure K, clicking on an individual sampling point yields a table of the most recent analyte readings from that single sampling point, Figure O. Clicking on the analyte name here yields a bar chart for the single sampling point, also analogous to Figure F.

Water Quality Measures:
Point-of-Entry: 126
Well: SP-830

Chlorine, mg/L	1.23	APR 19 2001
Coliform Bacteria	NEG	APR 21 2001
Fluoride, mg/L F	0.74	FEB 21 2001
Hardness, mg/L CaCO3	218.0	FEB 28 2001
Nitrate, mg/L N	2.3	FEB 21 2001
Sodium, mg/L Na	64.0	FEB 28 2001
TDS, mg/L	499.0	MAY 08 2000
THM, ug/L	0.0	APR 17 2001
pH, Std. Units	8.13	APR 19 2001

Figure O

3.6. System-wide Summaries

We maintain database tables for analytes across the entire distribution system, and we provide a means of a system-wide summary of water quality, Figure P. Clicking on a button on the main City Water screen (since it is applicable to the whole city, there is no map location to click on) opens a window summarizing all sampling points across the city into one average per analyte, representing the last 12 months of sampling. As with Figure M, the lower limit and upper limit indicate the sampling value at the 5th percentile and 95th percentile. The number of samples is of course substantially higher than in any individual zone.

Distributed Water Quality Summary

The information in this table is based on water quality monitoring results for the most recent 12-month period.

Water Quality Measure	Water Quality Measure Results			
	Average	Lower Limit	Upper Limit	Number of Samples
Chlorine, mg/L	0.84	0.45	1.24	3146
Coliform Bacteria, % Pos.	0.1	N/A	N/A	2962
Fluoride, mg/L F	0.42	0.14	1.0	750
Hardness, mg/L CaCO3	115	66	222	2449
Nitrate, mg/L N	1.81	0.41	5.4	750
Sodium, mg/L Na	41.0	19.4	69.5	2511
TDS, mg/L	264	162	493.8	3059
THM, ug/L	1.8	0.0	4.88	65
pH, Std. Units	7.82	7.32	8.21	3072

Figure P

Clicking on the analyte name in the distribution summary shows a bar graph analogous to Figure F and Figure N above, but representing the whole city instead of one water quality zone or one sampling point. Note that the city-wide summaries do not include the EMPACT points, because averaging in their data would overwhelm the much-less-frequent grab-sampling values.

Supply Source Water Quality Summary

The information in this table is based on water quality monitoring results for the most recent 16-month period.

Water Quality Measure	Tucson Supply Source				
	Clearwater	Avra Valley Wells	Santa Cruz Wells	Central Wells	South Side / TARP
Chlorine, mg/L	1.18	0.92	0.62	0.77	0.89
Fluoride, mg/L F	0.5	0.46	0.84	0.35	0.66
Hardness, mg/L CaCO3	84	79	191	117	235
Nitrate as Nitrogen, mg/L N	1.35	2.04	4.59	1.98	2.35
Sodium, mg/L Na	44	37	39	41	65
TDS, mg/L	230	In Progress	In Progress	715	In Progress
pH, Std. Units	8.13	8	7.71	7.82	7.94

Figure Q

Figure Q shows a water quality summary for the entire supply system, grouped by source (i.e., location of the well field). The "Clearwater" column indicates the water quality of the recharged Colorado River water, while the others indicate groundwater wellfields. Clicking on the analyte names here pops up a window with the analyte summaries, identical to Figure G; we don't present bar graphs for each supply source.

The data for the system-wide summaries comes from a database table called "sum_by_day," Figure R. This table is re-created from several sources in an overnight batch. It represents aggregated and filtered data in a ready-to-use form. For example, the daily average is pre-calculated per "locator" (either an SP or POE) and per "parm_syn" (analyte). Hence we can present a daily average by reading just one database record, which ensures a fast response time. To calculate monthly averages for the bar charts, we summarize 30 records (which means the monthly average is actually an average of 30 daily averages, as opposed to an average of the underlying individual values).

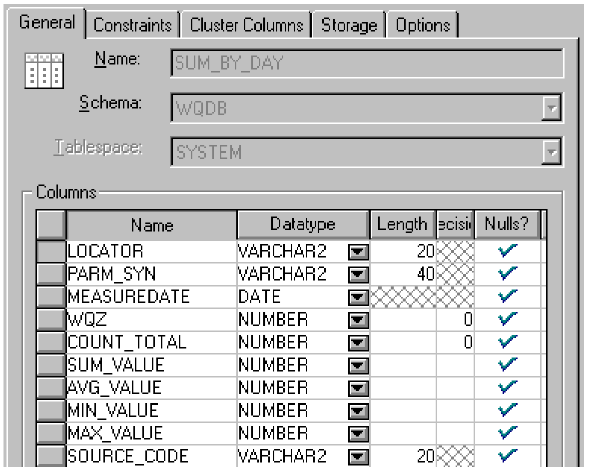


Figure R

The purpose of pre-summarizing the data is to allow for near-instant response even when analyzing large quantities of data. In particular, the percentile values (for upper limit and lower limit) are extremely slow to calculate (requiring an iterative process) and are pre-calculated in the "min_value" and "max_value" fields.

4. ISSUES

Prior to instituting a public web site for disseminating water quality information, Tucson Water used the more typical method of printing literature once a month summarizing the water quality for previous months. The informational brochures, while informative, were not particularly timely – there was sufficient time to “filter” the data, to review and aggregate it in a meaningful way, and to remove data which was inapplicable or difficult to explain to the public.

A public web site creates a situation where there is no opportunity to filter the data – we had to decide in advance how to aggregate and present the data. There is no mechanism to review water quality data before it appears on the public web site, especially for the automated EMPACT sensors, which record data and present it to the public with no human intervention. In the case of laboratory data and grab samples, people perform quality control, but the computer then automatically aggregates and selects the data – there is no human intervention beyond the individual lab technician who records the data.

The automated presentation of data enforces honesty in reporting – it is impossible to hide water quality problems at Tucson Water as long as the web site is functional. That level of honesty was one of the purposes of instituting this system: to maintain public confidence in the water system. Here we will discuss several issues arising from this method.

4.1. How can the public track bad water quality data?

Note on Figure E and Figure O that individual sampling points are reported in terms of “most recent values.” If a reading one day is out of compliance, it could potentially be hidden the next day by another reading which was in compliance. The aggregated data (bar charts) sum up all the values for one month, so an individual out-of-compliance reading might get lost in the averaging process.

The issue was raised that bad water quality data could therefore slip through the public’s notice. There are two responses to that complaint. First, the EMPACT points report daily data as well as the most recent value and monthly summaries (as in the line chart in Figure D). If this issue became politically important, the line charts could be extended to include other analytes. Second, members of the press could take the time to inspect the sampling points on a daily basis, and would therefore have access to every piece of water quality data, including any which were out of compliance. With some research into Tucson Water’s schedule for sampling points, the press could investigate sampling points as samples were taken.

This issue was raised further in the context of our presentation of percentile ranges instead of actually reporting the high and low values for each analyte (such as on Figure M). Showing the 5^{th} percentile and 95^{th} percentile values means that 10% of all water quality data is ignored, and the most extreme cases are discarded.

That concern is addressed by the same response as above – a diligent investigator could see the individual data behind the summaries. We chose the 5^{th} percentile and 95^{th} percentile ranges because they are the normal extreme ranges used for data reporting. Any analyst creating a paper version of the same data would perform much the same filtering method. Furthermore, the printed reports which were formerly Tucson Water’s only reporting to the public contained much higher levels of aggregation and filtering, which in this context means more than 10% of the data was ignored.

4.2. Why limit the number of analytes presented?

The Tucson Water web site reports on about a dozen analytes at about 450 sampling locations. The survey of sampling locations is very thorough, but the list of analytes is not – Tucson Water analyzes hundreds of analytes, as required by state and federal regulations.

We chose to limit the number of analytes presented in order to make the data understandable and accessible to the public. A long list of hundreds of chemical names would do more to hide relevant data than to present further publicly digestible information. Even with the short list of analytes that we did choose to present, press reviews concluded that "the data is well presented but there is too much of it" (paraphrasing television news reports from May 2001).

The most sensitive analyte omitted was E. Coli – we chose to report only total coliform. Our intent in this case was to avoid apparent duplication of data – i.e., any E. Coli problems would likely show up as total coliform problems, so we only reported the more general of the two. Since E. Coli is reported as "presence" or "absence" like total coliform, we concluded there was no meaningful quantification to reporting both. In the interest of brevity and accessibility, we chose to present just one.

Hardness is another issue of particular importance to Tucson residents. Typical homes in Tucson have a "dual-cool" system – an air conditioner for hot and muggy days and an evaporative cooler for hot and dry days. The evaporative cooler, or "swamp cooler," requires frequent filter changes due to calcium deposits, and many homes have water softeners for that purpose. Hence reporting hardness and other mineral content analytes is more important in Tucson than elsewhere.

4.3. Privacy concerns with individual properties

The map data on property parcels includes information on individual ownership (person's name and home address). We omitted that data from the map's parcel layers to respect individual privacy. Even though that same information is publicly available, we chose to avoid the ease of obtaining it via a web site, as opposed to taking a trip to the county assessor's office to look it up on paper.

We did originally show the street address for each parcel (the house number on each street) but removed that in the final version as well. We concluded that there was little additional value in presenting that information, since people would locate their homes by street intersections, and would recognize their own parcels without the additional numbering. The numbering itself just added clutter to the parcel screen.

We also considered allowing a "street address" lookup, e.g., one would key in "10 Main Street" and the map would zoom to that location and show the water quality for the nearest sampling point. We omitted that function primarily because of technical reasons unique to Tucson. First, Tucson Water does not service every address – there are competing water utilities in the city (Metro Water is the second largest after Tucson Water). We felt that if we linked to particular addresses, there would be an expectation that we reported the water quality for that particular address, which in the case of Metro Water customers, we had no means of doing.

Second, we felt that associating one sampling point with a keyed-in address would assign a relationship that was not accurate – any particular address gets its water from several sources. We concluded it was better to let viewers navigate themselves to nearby sampling points rather than choose one to associate with an address.

In the wake of Sept. 11, there is some concern over the presentation of exact locations of sampling points, because saboteurs could readily locate points-of-entry in the distribution system. Of course, we designed and implemented this system before Sept. 11. We will perhaps reconsider the display of sampling locations once federal recommendations about water infrastructure security are concluded.

5. CONCLUSIONS

One author (JAG) uses the "Wanda Indicator" to determine public concerns about water quality. Wanda is a long-time resident of Tucson who has no interest in Tucson Water except as a customer and one who drinks the water. If Wanda is concerned about an issue, it indicates that the issue has reached the concern of the general population. The author calls Wanda every several months to ask her what she thinks of her water, and what she has read in the paper. At this writing, Wanda drinks Tucson tap water, albeit with a home filtering mechanism because she is still concerned about water quality. One of her neighbors drinks unfiltered tap water. Perhaps that is our best measure of success.

A map-based system for presenting water quality data has assisted in a public relations success with Tucson Water's new water source. The situation in Tucson required a substantial investment in public relations because of the switchover to the new water source. The tools developed for Tucson Water are applicable to other water systems across the country, providing for public dissemination of large quantities of information in a readily-available manner.

The EMPACT grant for this work was one of the last EMPACT grants issued by EPA. EMPACT grants were intended to disseminate information to the public about their water supply and other topics. The current administration has canceled the EMPACT program.

This paper is available on-line at http://www.drinkingH2O.com/imtech/ .