#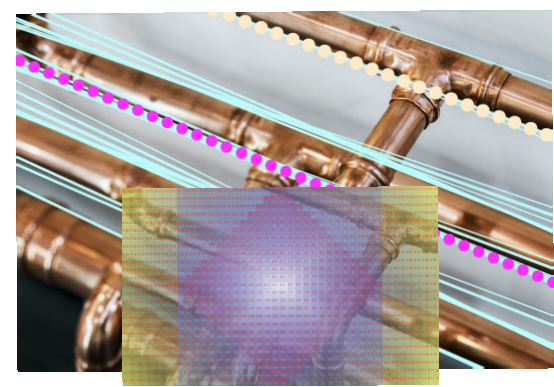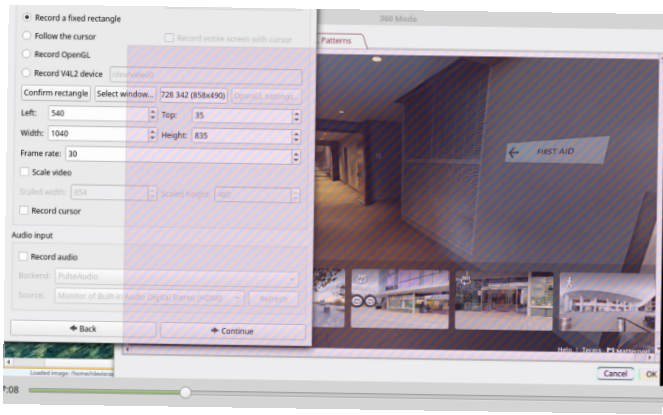 TorqA3R: Cloud Hosting Architecture and Web/Desktop Hybrid Software Engineering, as applied to Multimedia Government, Civil and Environmental Informatics

Amy Neustein, Ph.D.
CEO and Founder
Linguistic Technology Systems
Fort Lee, NJ
917-817-2184
amyneustein@verizon.net

A New Cloud/Native Hybrid Software Engineering Framework for Embedded Web Applications, Government Information Services, Urban/Community Development, and Environmental Technology

Introduction

Recently-developed technologies such as Cloud-Native hosting services, Embedded Web Sites, and Web/Desktop hybrid software applications present currently-untapped potential for next-generation civil and governmental information services — particularly in the context of sustainability, Civil Engineering Informatics (**CEI**), community development, and digital government.

Linguistic Technology Systems (LTS), a startup in Fort Lee, has engineered software components supporting next-generation information services with an emphasis on Translational Informatics — a term used to characterize how science and research can be "translated" to benefits for individuals and communities, often abbreviated as **TRI** (for "Translational Research Informatics"). We would like host governmental and/or civil data sets or multimedia resource archives which leverage and demonstrate this new technology. LTS can set up the cloud-hosting infrastructure required for running services with the unique features described below. For reasons explained below, most cloud hosts do not support the kind of Cloud-Native technology described here; as a result, a publicly-accessible cloud environment explicitly designed for such technology (with precompiled libraries and similar programming utilities) could be a beneficial public service, enabling developers to run software on the cloud that would be very difficult to deploy otherwise.

TRI and "Translational Science" was originally formulated in the context of bioinformatics and medicine, but more recently has been applied to other fields, including those related to climate change, environmental health, and public policy. We believe that **TRI** protocols initially developed for tracking the progress of biomedical and pharmacological science "from bench to bedside" — i.e., from the laboratory to beneficial clinical interventions — can be adapted toward the documentation and analysis of scientific initiatives which, often through government actions, are operationalized toward community development and helping the environment. Proper technology can support greater transparency and accountability for both government and science. To this end, we have worked on a **TRI** data-sharing and software engineering framework that we call **TorqA3R** (Translational Research Query Protocol – Application as a Resource) with potential applications to civil/governmental data management.

In particular, we would like to develop a web service that could supplement and integrate governmental and environmental data for Bergen County and/or New Jersey as a whole, with an emphasis on environmental health, sustainable design, and civil engineering. We propose to engineer this service with a focus on supporting web-site embedding into standalone, desktop-style computer applications, which is a relatively new way of accessing web resources. Embedded web sites are used in specially designed standalone applications rather than visited exclusively via web browsers (although conventional browser-based access is still possible). Web-site embedding allows web services to be enhanced with desktop-style technology that would not ordinarily be available browser-based web applications, such as showing multiple, interconnected multimedia windows simultaneously, or presenting isolated windows with structured data that complements content seen in a primary window — for example, presenting details about a building or construction project in a dialog box appearing alongside representations of that building/site in a video, street view tour, or street map.

Code implementing **TorqA3R** web sites/applications or services can be packaged and shared in a portable manner, so that the site may be run on multiple computers. Simultaneously, **TorqA3R** services are optimized to be accessed from self-contained, desktop-style software, using dedicated windows which render web pages essentially as miniature browsers, via technologies such as **QWebEngine** — which functions as an instance of the Chrome browser local to a single (self-contained) application. When embedded in this manner, web sites/services can benefit from more sophisticated graphics, visualization and multimedia capabilities, exposed through the host application, than are possible through conventional (non-self-contained) browsers insofar as these expose a narrower range of functionality to intra-web application code.
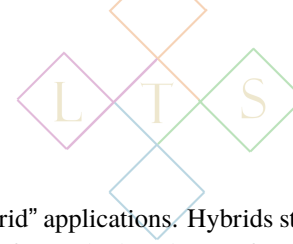
In general, sites which employ both browser-based and native functionality are known as "hybrid" applications. Hybrids straddle the line between web and desktop paradigms. Engineering hybrid applications requires a software-development platform which is distinct from both the mainstream native frameworks (such as **QT** or **MFC** — Microsoft Foundation Classes) as well as the popular web toolsets (**Node**, **Rails**, **Django**, and so forth).

At the same time, once a shared instance of portable/multi-peer web applications becomes hosted in a public location (on the cloud, for instance), users can access **TorqA3R** services without a local version of the relevant application on hand. This includes both members of the general public who prefer not to download their own copy of the underlying software, as well as users who do have the software locally installed but will sometimes access the shared content from locations where their local program is unavailable. The **TorqA3R** model supports private data spaces where users can access their own specific content via a cloud share-point as if they were working from their own personal copy of the web service.

Likewise, common share-points can host data sets and resources that are too large for individual computers, such as **GIS** tile servers (storing gigabyte-scale collections of digital-cartographic images, to create street maps and other **GIS** rendering elements) and sizable arrays of video/multimedia files. Similarly, functionality that is impractical to implement in a cross-platform manner may be encapsulated in a virtual application image (via technologies such as **docker**) functioning as a local web server, so that native front-ends are driven by back ends with their own virtual operating system. Apart from use-cases where such virtualization is unavoidable, however, **TorqA3R** could run as wholly native applications with the entire code base on a local computer, while simultaneously deploying a copy of the relevant server-side components on the cloud.

Overall, **TorqA3R** combines several comparatively new technologies, such as **QWebChannel** or equivalent bridges to marshal data between web and native components; viewers (and parsers for metadata associated with) multimedia assets, e.g., 360° photography or tile servers with detailed **GIS** layers that can customize map-segment images to individual projects/applications; and "cloud-native" containerization for self-contained, native-compiled software instances. Perhaps because of these constructs' relative novelty (5-10 years, as compared to the 20-25 year span by now of "Web 2.0") the infrastructure for supporting shared applications according to the criteria described here is not commonly available, by contrast (for instance) to ordinary cloud-native architecture supporting browser-based rather than embedded web sites.

For this reason, we encourage government initiatives which leverage hybrid-application and web-site embedding for next-generation civic/administrative information services — with the added side-benefit of concretizing an infrastructure that could also be used by the public at large. For example, cloud-hosting services with explicit support for hybrid/embedded web applications can be a publish-point for government web sites but, in so doing, space might be reserved for other kinds of applications maintained by private individuals or institutions, as an alternative to corporate cloud providers (e.g., Amazon Web Services, or **AWS**). Similarly, resources such as **GIS** tile servers or government databases could be made available to the general public for a diverse range of use-cases (digital maps, for instance, benefit from customizable tile-server providers as an alternative to commercial programs, such as Google Street Maps, whose **GIS** features are limited and occluded by practices that support their commercial business model; e.g., showcasing enterprises like stores or restaurants on map displays, whether or not these are relevant to the software where the maps are viewed).

Web-Site Embedding and Multimedia

Web-site embedding is particularly useful in the context of multimedia databases where different media formats can be interconnected; e.g., geotagging video frames; linking **GIS** map locations to videos or Virtual Tours of sites and buildings; or cross-referencing specific page-coordinate locations in **PDF** files with map locations or with multimedia files. Interactive cross-referencing provides enhanced, visually informative user experience, such as seeing urban/environmental redevelopment projects via **GIS** maps, videos, Virtual Tours, and image-series in combination; however, this sort of user experience can only be achieved in a limited fashion in the browser. Most web sites, however — and, in fact, almost all web-development platforms — are designed exclusively for browser-based access, and do not make any special provisions for use as embedded resources. To address these lacunae, we have engineered protocols to standardize how web sites may interoperate with host applications, allowing both sides to maximize web-site embedding's benefits. Consequently, we would like to propose the curation of a government, civil engineering, sustainability, and environmental data web service that would implement these protocols when used as embedded resources, while also being accessible as ordinary browser-based applications.

In addition to superior user experience, web-site embedding allows host applications to unify several isolated web sites which are logically (but not logistically) interrelated. This is particularly appropriate for governmental or environmental web sites, that tend to be artificially separated by jurisdictional boundaries. For example, a researcher or governmental official interested in the environmental impact of municipal decisions — such as sponsoring public/private initiatives or zoning and land-use classifications — will find relevant data scattered across multiple websites partitioned by geographic boundaries. New York City, other counties in the State of New York, and counties in Northern New Jersey each have distinct websites and their own offices for requesting information. Maps such as NYCPlanning's **ZOLA** system, which provides zoning and property info for the five boroughs, are blank across the river in New Jersey; while analogous resources such as New Jersey's Department of Environmental Protection (**NJ-DEP**) interactive maps contain no information back in New York.

And yet, most environmental risk factors are shared between the states of New York and New Jersey — hurricanes, flooding via the Hudson River or Atlantic Ocean, poor air quality, and so on — and likewise housing and quality-of-life trends. Unfortunately, however, because state agencies on opposite sides of the Hudson River operate separately from one another (and answer to disparate governments; Albany there, Trenton here) there is little incentive for New Jersey and New York State digital resources/platforms to be synchronized or interoperate.

Moreover, in the same way that useful data is often siloed in separate state and/or county web sites, important links between different kinds of multimedia content can be similarly neglected. For example, maps, videos, image series, and virtual tours can all be used to showcase environmental restoration and/or urban-development projects, but most software applications have no way to connect different files which are interrelated — for instance, connecting videos of a new construction project to interactive virtual tours of the same buildings. The LTS software for **TorqA3R** and "Translational Sustainability Informatics" (see below) addresses this problem by establishing a protocol for multimedia cross-referencing and providing software which can display diverse media types and follow cross-references so that users can browse between multiple interrelated file types.

Here are several examples of logical interrelationships that are meaningful when integrating diverse forms of multimedia in **AEC** and **CEI** contexts:

Building Inspections Companies which do residential and/or commercial building inspections — including several in New Jersey — have started to offer 360-capture photography as part of the inspection process. 360° photography produces virtual tours that can expedite initial or follow-up inspections, and document the history of a property's inspections, permits, and licencing. Embedding virtual tours in desktop-style applications allows individual locations inside the buildings to be synced with government records, so users viewing the tour can seamlessly transition to accessing an inspections database (so as to research, for example, a commercial property's compliance with facilities codes, or the status of renovations).

Geotagging Videos Individual video-frames — and in general the portions of a video running between two time-stamps — can be annotated to declare that the video (at a given moment, or for a specified period of time) depicts or documents a real estate property or geospatial location with specific **GIS** coordinates. People watching these videos can then benefit from viewing them alongside interactive **GIS** maps. Consider videos documenting environmental cleanup projects and rehabilitation of degraded land areas: while seeing the progress of a construction site, restoration of waterways, development of parts and other greenspace, etc., users can benefit from observing their positions on a map and noting relationships to surrounding streets, waterways, topography, and so forth. Special technology, however, is necessary to extract video-frame annotations and pass embedded **GIS** data and coordinates to geospatial displays — and, in general, to develop computer software where **GIS** maps and video players co-exist in one application.

Converting street-view or virtual tours to video Virtual tours allow users to select their own (digital) path through a building or outdoor site, but for some purposes it can be helpful to run a tour with a specific path selected for demonstration and documentation purposes, and then capture the results as a video that may be replayed with no changes to the walk-through (such video-capture also requires special-purpose software). Street-view tours of an urban neighborhood can likewise be converted to videos, which can be useful for similar reasons.

Emergency Management and Hazard Mitigation Cross-referencing digital maps with videos, photo galleries, and/or virtual tours can document how jurisdictions identify and mitigate environmental and/or public risk factors. For example, maps and videos (including those derived from street views) might jointly show flood or fire evacuation routes; identify industrial sites where there exists a contamination risk, or similar public health concerns; document patterns of erosion, drought, floods, soil change, and similar indicators of climate change having an effect on local ecosystems; and track potential ecological antagonists such as invasive or harmful species (e.g., toxic algae).

Government Services Syncing maps with government data sets can make it easier to provide information about public services to residents of (or visitors to) local communities. For example, a single portal which loads public transit data alongside maps and street views can help phone operators direct customers to their correct bus stop, in locations where there are several nearby stops serving different routes; or where customers need stops with or nearby specific facilities, such as a seating area, a location to buy tickets, or a convenient waiting place (e.g., a restaurant or coffee bar).

Zoning and Land Use Adapting digital maps as an entry-point for obtaining information about zoning, land use, residential/commercial property data, local ordinances, and so forth, can expedite public use of government databases. For example, Zoning and Land Use maps could be linked with photos, videos, 360° photo assets, or **PDF** files which visually and/or interactively explain the restrictions/requirements enforced for locations within different land-use areas. Relevant documentation could then be loaded directly from **GIS** coordinates based on the zone classification of a selected **GIS** location. In the case of **ZOLA**, for example, New York City residents can open a small box summarizing the definition of the applicable land-use designation for any map location, but further details about the relevant zoning regulations are only available from a separate part of the web site. A more multimedia-oriented front end (using an embedded web site) can load a **PDF** document or a video explaining the local zoning parameters directly from the digital map, superimposing it over the map window near the specified **GIS** coordinates.

Track Social and Development Trends Embedding interactive maps with data sets that show timelines or historical data about trends such as population changes, urban renewal projects, patterns of construction and community development, or impacts of environmental policy, can help to document and assess the influence of municipal, county, or state laws/regulations and initiatives on environmental well-being and quality of life. In particular, insofar as jurisdictions might cite specific (quantifiable) metrics for government policy — acknowledging parameters such as optimal population density, housing availability, low carbon-footprint construction, reductions in air/water pollution, improved traffic flow, etc. — multimedia data sets and presentations can document the effectiveness of programs or regulations at meeting their declared goals. It is one thing, say, to state in words a jurisdictions' commitment to ecological cleanup/restoration or the construction of new residential properties; but a more convincing presentation will actually document the sites and locales affected by such policies with pictures, videos, or virtual tours. Indeed, quantifying the effectiveness of policies adopted for ecological and/or community-oriented reasons serves an important role of clarifying how well scientific models of environmental health and sustainability are *translated* to concrete societal action, continuing the analogy with "translational medicine."

We believe that governments, researchers, and the general public deserve an integrated portal which aggregates data about climate change, environmental factors in public health, and sustainable engineering and design solutions with demonstrable social and ecological benefits. Data integration is especially important because health, environment, energy, and civil problems and solutions are tightly interconnected. For example, problems exacerbated by global warming — such as toxic algae blooms — cause ecological damage (polluted waterways) that in turn yield public health risks (cyanobacteria poisoning) and economic harm (via agricultural damage). Likewise, local government policy (such as zoning and land use classification) shape population density and traffic flows, which in turn affects air quality, energy consumption, and carbon emissions. At the same time, sustainability initiatives — such as Superfund designations — have positive impacts that branch in multiple directions, including better environmental health (both through removing pollutants and transitioning former industrial areas to natural, potentially carbon-sequestering habitats) and social/economic benefits as superfund sites are redeveloped (parks and natural areas for quality of life, and/or residential or commercial areas for economic benefit). These are a few examples of converging energy/environment/ecology/engineering/economic factors.

Consider the following scenario: someone intends to study land-use patterns in the context of global warming and natural hazards; specifically, how to formalize the assessment of hazard risk relative to different zoning classifications (industrial, commercial, residential, with different population-density averages). There are several data sources which would be logical places to acquire relevant information, including the **ZOLA**, **NJ-DEP**, and the World Food Programme **PRISM** "climate risk monitoring" project, which receives Earth Observation data and cross-references it with "information on socioeconomic vulnerability ... to inform disaster risk reduction and social assistance programmes." In this hypothetical example — merging land-use and hazard-risk data for both states via the **PRISM** system — the three specific components involved (**PRISM**, **ZOLA**, and **NJ-DEP**), although each operate via an interactive **GIS** interface, are technologically incompatible; there is no way to merge data from one map onto either of its peers. The **ZOLA** display shows no land-use data for New Jersey, and vice versa for **NJ-DEP** vis-à-vis New York; and neither map can show **PRISM** data layers

In general, information curated at local, state, and government levels which pertains to collective environment/engineering/economic factors tends to

be piecemeal and fragmented, failing to recognize the factors' interconnections. Data which is overlapping and jointly relevant tends to be split among disparate web sites and databases with no cross-referencing, data-discover mechanisms, or acknowledgement of important inter-relationships. To illustrate this problem in concrete terms, here is an example of several governmental and environmental resources specific to Northern New Jersey:

✦ New Jersey Department of Environmental Protection: **NJ-DEP** provides a collection of interactive maps which support visualizing and querying various data sets related to environmental protection, wildlife, agriculture, pollution, and related ecological details.

✦ NJ-MAP, developed by Rowan University, provides interactive maps covering a range of domains across both natural and urban/governmental data sources, such as New Jersey Conservation Blueprint (resources specific to protecting natural land areas) and a Parcel Explorer that aggregates geographic and land-use data for individual property lots.

✦ NJADAPT, part of the Climate Change Resources Center at Rutgers University, provides maps and data-visualization for studying the effects of climate change on New Jersey communities, addressing both natural and man-made ecosystems.

✦ New Jersey Division of Water Monitoring & Standards (**DWM&S**), in conjunction with Rutgers University, provides interactive maps with a "Continuous Data Monitoring Program" yielding a frequently-updated publicly-accessible portal specific to hydro-ecology.

✦ County-Level Offices of Emergency Management (**OEM**), as well as the state-level **OEM**, are responsible for curating and sharing information about natural and man-made emergencies. This includes a **FEMA**-mandated Hazard Mitigation Plan, typically developed individually within each county with data aggregated from each jurisdiction, that requires local governments to document the steps they are enacting to help prevent loss of life or property damage in the event of natural disasters and other hazards.

✦ The National Institutes of Environmental Health Sciences (**NIEHS**) provides a suite of tools for publishing information connecting environmental and public health, ensuring that data from disparate sources becomes interoperable and collectively analyzable. This includes a "translational research" framework organized in concentric "rings" progressing outward from biomedical/biochemical investigations to public health and environmental policy. Likewise, **NIEHS** established an "Environmental Health Language Collaborative" to define a common format and semantics for asserting environmental and public health observations and data sets, including those relevant to describing environmental hazards and mitigative steps such as those addressed by Hazard Mitigation Plans.

✦ **ZoLa**, New York City's Zoning and Land Use information service (mentioned earlier) was implemented on top of an innovative code library for **GIS** web services, called Ember Mapbox Composer, that itself was developed and published by NYCPlanning. Although **ZoLa** is specific to New York City, Ember Mapbox Composer is independent of any particular database or jurisdiction, and could be reused by other cities and states.

✦ **PRiSM** (Precision Manufactured Housing), a project developed by the city of London (UK), is a tool to spearhead sustainable **AEC** practices in the London metro area (note that this **PRiSM** is unrelated to the **PRISM** portal, referenced above, curated by the World Food Programme). **PRiSM** tools include **GIS** and **CAD** displays as well as charts/plots and other data-visualization resources relating to civic **AEC** data in the London region. Although specific to London (the city did not provide a generic back-end analogous to Ember Mapbox Composer as the technology behind **ZoLa**) it is probable that other cities could benefit from applications modeled around **PRiSM** (perhaps multiple cities could collaborate to create a version of **PRiSM** not tied to London or any other particular geographic location).

✦ The other **PRISM** — the one implemented by the World Food Programme (**WFP**) – is also a useful resource for risk/hazard mitigation due to environmental factors. The backstory behind this **PRISM**, insofar as it was spearheaded by an organization whose stated purpose involves combating global hunger instead of climate change, was the **WFP**'s recognition that climate change exacerbates the socioeconomic stress factors which contribute to nutritional insecurity. The organization therefore built **PRISM** with the explicit purpose of pairing socioeconomic data with environmental observations, using domain-specific techniques to extract socioeconomically salient data sets and trendlines from Earth Science observations (correlating droughts with food shortages, for example, or quantifying how climate change affects internally displaced persons and emigration). **PRISM**, in turn, receives environmental data from the Committee on Earth Observation Satellites (**CEOS**) via Open Data Cube (**ODC**), a platform with multiple domain-specific instances, including the **WFP**'s "Humanitarian Data Cube." Other initiatives, such as the Climate Hazards Center (**CHC**) at the University of California – Santa Barbara, employ a similar analytic framework, and the Humanitarian Data Exchange project, via Humanitarian Exchange Language (**HXL**), offers tools for merging thematically connected data sets such as those from **WFP** and **CHC** to augment each other's scope and detail.

✦ The Institute for Sustainable Infrastructure (**ISI**) "envision" framework models 64 "sustainability and resilience indicators," also called "credits," which document sustainable practices adopted per-project by entities in the Architecture, Engineering, and Construction (**AEC**) sectors. The World Resources Institute (**WRI**) "**ZCB** Pathways" framework is a similar method for tracking progress toward **ZCB** (Zero-Carbon Building) emissions goals. By applying the **ISI** and **WRI** models to new or existing buildings in their communities, local and state governments can document how they are addressing climate-change related risks by reducing buildings' energy use and carbon footprint. The significant detail about **ISI** and **WRI** models is that, in additional to generic "best practice" recommendations, they have also formulated specific metrics, assessment tools, and iconography which quickly documents how extensively and in what manner buildings adopt sustainable practices, making it possible to aggregate data from multiple properties into a more holistic model of trends in the context of specific locales and communities. For instance, **ISI** or **WRI** icons can be employed in **GIS** maps; a higher concentration of icons in given localities indicates that **AEC** firms in that area (perhaps with influence from local governments, via leveraging tactics such as land-use abatements or public/private partnerships) have embraced sustainable practices within multiple sites and projects.

The concerns addressed by each of these projects or organizations are overlapping and in many ways mutually reinforcing. For example:

1. Identifying environmental risk factors overlaps with emergency management, insofar as general threats due to ecological damage can transition to emergencies demanding immediate responses in the wake of natural events or human accidents. For example, polluted waterways (due to toxic algae, as mentioned earlier, or due to industrial chemicals) that present a baseline threaten to human or animal health in their vicinity can become more hazardous due to storms or floods that cause the waters in affected areas to spill over into rivers, aquifers, and other usually safe waterways. Similarly, industrial sites containing dangerous materials present a public health hazard in the event of forest fires. Data for tracking sites with long-term environmental-health consequences will intrinsically, therefore, overlap with formulating protocols for occasional emergency management. Conversely, improving sites presenting long-lasting pollution threats is an effective form of hazard mitigation and risk reduction/management.

2. Sustainable architecture and engineering categories developed by the **WRI** and **ISI** are associated with specific buildings, land-areas, and construction projects. Both **WRI** and **ISI** offer a series of terms and icons that indicate sustainable practices adopted at specific locations; because these indications are tied to geographical locations, they can be incorporated into **GIS** maps addressing environmental and climate-change factors. Mapping the adoption of sustainable practices can help to visualize trends in where such practices are being applied more thoroughly (which communities are especially forward-looking in this regard) and can document relationships between regions that embrace (or lag behind vis-à-vis) sustainable **AEC** and areas projected to be most affected by climate change (flooding, forest fires, rising ocean levels, etc.)

3. Zoning and land-use classifications, which are typically the responsibility and under control of local jurisdictions, have a significant effect on the climate footprint of urban and suburban communities — affecting traffic patterns, concentrations of population density, and environmental impact of industrial development, plus the consequences of clearing natural land for construction in the first place. Municipalities may choose to alter zoning policies either holistically or on a case-by-case basis, under the influence of environmental impact studies; in some cases governments might offer abatements for individual developers who propose projects with direct social/environmental benefits — such as well-situated residential units — or eco-friendly enhancements, such as commercial developments paired with greenspace-preserving parks or publish-use areas. Digital records can document zoning and land-use decisions and proposals, along with relevant data and analyses, adding greater transparency to this process and also encouraging neighboring districts to work in consort toward sustainable Civil Engineering practices.

4. Multi-Media technology is important for presenting sustainable **AEC** solutions. Sustainability in Architecture and Construction addresses several facets of building design, including physical materials, cost-effective construction techniques, and energy efficiency post-occupancy. Organizations such as **ISI**, **WRI**, and the International Institute for Sustainable Development (**IISD**) — alongside localized projects such as London's **PRiSM** software — provide **AEC** recommendations across these design areas. However, such documentation is often cursory and hard to visualize; for example, it is difficult to find videos or virtual tours showing sustainable-engineered buildings from a practical perspective. In the case of Precision Manufactured Housing (**PMH**), say — which is a promising new technology for cost-effect **AEC** implementation — it is difficult to visualize what **PMH** buildings look like, what the experience is like to live or work in them, and so forth. Similar comments could be made about "smart building" techniques using designs and materials (such as natural cooling) to reduce carbon footprint. Multi-Media presentations can help by syncing image series, videos, and virtual tours with **2D** floor plans or **3D CAD** presentations: having the opportunity to see sustainable **AEC** techniques formulated within **CAD** software and then transition to seeing the real-life results, via video tours showcasing buildings, or virtual building walk-throughs, with sustainable design features and materials highlighted via annotations and/or supplemental documentation. In the case of modular **PMH** design, for instance, what are the individual building parts; how do they look pre-assembly versus on-site; how do modules piece together into an organically holistic design; and how can "smart" features be showcased insofar they are installed and utilized in a completed building?

5. Eco-friendly **AEC** practices need to be evaluated in conjunction with sustainable land use patterns. "Green" **AEC** can promote sustainably-sourced and/or energy-efficient buildings from a site-specific point of view, but **AEC** environmental impact is also more holistic. For example, construction projects which disrupt natural habitats and/or are poorly connected to local neighborhoods and transit routes (causing, e.g., longer car commutes to nearby cities) can have a net negative impact even if the buildings themselves are designed in an eco-friendly manner. The domain of "sustainable" architecture includes using building materials and designs which minimize environmental footprints both when sourcing construction and post-occupancy, but this domain also involves construction protocols that offset the exceptionally high **AEC** costs typical of US urban areas. In the best-case scenarios, new buildings will simultaneously optimize sustainable design, construction cost relative to the number of residential units created, and integration with local communities and transit options (e.g., siting eco-friendly buildings on underused lots close to urban centers is more eco-friendly than comparable exurban projects). Correctly analyzing these multiple optimizations factors requires cross-domain data integration.

6. Government incentives, such as Superfund designations and public/private collaborations, are generally assessed for both their environmental and economic benefits. Therefore, metrics for defining project goals, and subsequently performance evaluations, call for quantification of both environmental and economic impacts, as well as integrated information sources providing the requisite data in each domain. From the public's perspective, for example, **GIS** services mapping out Superfund and/or public-private codevelopment sites can link each site to documentation of the commercial entities that stand to benefit from the projects (and that are accountable for ensuring that projects actually show their anticipated positive outcomes) as well as environmental-impact projections.

As these examples suggest, sources which provide information within one specific domain (such as **AEC**, environmental data, zoning and land use, Civil Engineering, and so forth) are most valuable when they can be cross-referenced with interrelated data addressing parallel concerns, linked by **GIS** coordinates, municipality, or initiatives (for instance, multiple Superfund sites across Bergen, Hudson, and Passaic counties; or Superfund locations in one borough linked to other climate-mitigation projects within the same jurisdiction). Our proposed dashboard and database would operationalize such data-integration by unifying multiple data-sources both at the User Interface level and via a multi-source data query mechanism.

In general, this kind of data synthesis proceeds along two parallel tracks, source-integration and application-integration. Application Integration means providing one single application which allows users access to multiple data sources. For this proposal, we envision a single self-contained desktop-style application that would embed websites such as those listed earlier (and any other ones which are relevant) as well as free-standing map views, images, charts/data-visualization, videos, 360-photography tours, and other multimedia assets. Users could switch between multiple data sources or view them side-by-side. To support multi-source interoperability, the host application would define a common protocol for user interactions with each of the interconnected web sites and data back-ends. Data-Source Integration, conversely, involves constructing or stacking **API**s, **ABI**s, and other integration layers to ensure that disparate data sources can be accessed via common data models, access protocols, and query systems. Wherever possible, such integration would extend existing sources' capabilities (such as **API**s) to honor Common Data Models and other type- or query-level protocols designed to merge multiple source-points into a common information space. When directly modifying data-sources is not possible, adaptation layers can be introduced to translate common protocols into **API**/query requests that the underlying data source can recognize.

Translational Sustainability Informatics

Taking inspiration from "bench-to-bedside" research in Translational Medicine — referring to the "lab bench" and in general the steps requisite for translating biology and bioinformatics to concrete clinical practice — Translational Sustainability encompasses tools and perspectives needed to lever-

age contemporary science toward sustainable and eco-friendly human communities. In the context of computer software, Translational Sustainability emphasizes tools related to **AEC** and **CEI**, particularly towards the goals of designing "green" residential, commercial, and industrial buildings; environmental cleanup; and successfully integrating social and environmental concerns via urban planning (increasing housing density while decreasing car traffic, for example, or creating parks and nature-preserving spaces that benefit the environment as well as public health/quality of life).

The term "Translational Sustainability" is sometimes used (e.g., by a University of Michigan research center) to describe investigations into translating scientific results in fields such as ecology, chemistry, and materials science into practical uses for eco-friendly and sustainable **AEC**. This is one of several recent scientific disciplines described as "translational" by analogy to clinical medicine. Analogously, "Translational Informatics" addresses the overlap of computer science with translational research in disparate domains, with respect to data warehousing, integration, and analysis — again inspired by biomedical practice (bioinformatics) but generalizing to science in general, alongside **CEI**, geoinformatics, cheminformatics, etc.

Our interest in Translational Sustainability is focused on using multimedia software engineering to plan and analyze sustainability initiatives. More generally, multimedia software improves the overall ecosystem of civic data managed by governments at different levels, from individual towns and cities upward. We are committed to developing software which synthesizes and cross-references different kinds of multimedia content, in contrast to most current software applications, which will typically handle only one or two media types. Integrated multimedia software allows users to access diverse multimedia content, and multiple embedded websites, at one time, which is more convenient from switching between applications and/or websites even if their content is inter-related.

We suggest the hybrid phrase "Translational Sustainability Informatics" (which could be abbreviated as "**TSI**") to express the intersection of "Translational Sustainability" and "Translational Informatics" — focusing on data management and Software Engineering in the Translational Sustainability context. **TSI** can combine implementations in several areas, particularly these:

1. Environmental Science: Assessing how measurements of environmental health/degradation and climate change — plus simulations of future climate conditions and the climatological effects of human activities and man-made materials — can inform proper steps for mitigating our deleterious effects on the environment. For example, ecological data can document trends before and after climate-mitigation interventions.
2. Health Science: Documenting effects of toxins, pollution, and climate-related hazards (such as algae blooms) to human and ecosystem health via biomedical technologies. The **NIEHS**, for example, is a subdivision of the National Institute of Health; much of the former's research addresses modeling the effects of ambient toxins on cells and organ systems, and measuring or simulating such effects via bioimaging, in-vivo/in-vitro experiments, digital reconstructions, and so forth. Epidemiological data can also track environmental effects on human populations as a public health concern.
3. Architecture, Engineering, Construction, and Sustainability: Exploring how fields such as materials science, architectural physics, energy-efficiency, and industrial manufacturing can yield practical **AEC** tools which optimize building's cost-effectiveness both in terms of progressing through planning, design, and construction phases and post-occupancy. Examples questions include how to apply modular/prefab or **DfMA** (Design for Manufacturing and Assembly) techniques to reduce construction costs, and how to achieve natural cooling, ventilation, shade, solar heating, and other organic connections with buildings' environment to mitigate the need for electricity to create comfortable living conditions.
4. Civil Engineering and Sociodemographics: **TSI** can overlap with Civil Engineering Informatics (**CEI**) and related information-management strategies, particularly in the context of governance and urban development. Analyzing population trends and socioeconomic data, for example, can clarify the relationship between urban communities and their surrounding natural ecosystems. Environmental conservation strategies will necessarily be different in medium-to-high density urban areas — with small land-areas dedicated to greenspace that are integrated into residential and civic infrastructure — compared to more distant rural and exurban zones. Societal trendlines can also impact connections across the boundary between human and natural environments: for instance, a shift toward remote office work (and away from commuting) can have the positive benefits of less car traffic, but it also points to concerns about sustainable infrastructure closer to where workers live — widening the area where eco-friendly and community-oriented civic development strategies are important outward from office-centered downtowns to cover large swaths of metro areas.

Like many "informatics" disciplines, **TSI** can be focused on formalizing canonical representations for data sharing and reuse. Bioinformatics, as an example, often concentrates on standardized encoding for genomic data (gene sequences, transcriptomes, mutation-evolution) and preserving/annotating diagnostic images; canonical genetic and picture-archive formats ensure that information can be shared across many different access points (multiple hospitals, research centers, diagnostic labs, etc.). Similarly, geoinformatics is structured around the common reference point of geographic coordinates, through which any data with geospatial components can be cross-referenced — most familiarly via latitude and longitude, but **GIS** systems also often use other **GIS** units, such as those based on tile-map renderings or Mercatur projections. Geoinformation technology needs to properly annotate information source to indicate which coordinate systems are employed for any given data set (and also provide algorithms for inter-converting between systems, e.g. latitude/longitude to Mercatur formulae). **TSI** can similarly focus on standardized data-representation techniques, coordinate systems, units of measurement, dimensional analysis, and related protocols to ensure interoperability between data sources representing the distinct scientific concentrations feeding in to Sustainability research (ecology, sociology, engineering).

As with other informatics branches, the proper role for **TSI** involves best practices formulated for reusable software and application packages — in short, code models and protocols alongside data models. For one concrete example, consider sustainability documentation in the context of modular architectural design. Controlled vocabularies, such as those proposed by **ISI** Envision and the **WRI** Zero-Carbon framework, or similarly the **NIEHS** Environmental Health Language Collaborative, define standards which software should adopt when serializing and describing data for sharing with other software components. However, similar semantic requirements come to the fore in the context of software application design in the first place. There are, for example, plugins to **CAD** programs which are used to design modular-housing units; such plugins allow **PMH** (precision-manufactured housing) or **DfMA** data models to be recognized and visualized within **CAD** front-ends, and incorporated into calculations alongside other **BIM** (Building Information Management) and **IFC** (Industry Foundation Classes) profiles. This is not only a data-sharing issue, but also affects the **CAD** User Experience directly; the design screens for **PMH** units, say, would be enhanced to incorporate user-visible icons and options for loading **TSI**-specific data-entry/summary forms or algorithms/computations (for example, calculating cost-per-unit for **PMH** manufactured components, or visualizing how multiple modular component are interconnected in a completed building).

In short, we envision an application-oriented **TSI** information model that defines protocols for managing **TSI** data within user-visible applications (including **CAD** programs, blueprint viewers, property-management software, and Civic Data engines such as **PRiSM**) as well as inter-application data

sharing operating more behind-the-scenes (i.e., typically not user-visible). The **TSI** protocols in this sense can address such concerns as systematic description of user-generated events when viewing **TSI** data and/or **UI** components; populating **UI** components with **TSI**-related data types; icons and terminology for identifying **TSI**-specific data and operations in the context of software which is not specific to **TSI** (e.g., **CAD** and **AEC** technology); and standards for representing **UI** state when sharing data across applications.

PDF/SVG Formats and Text Mining

Earlier comments about multimedia cross-referencing also apply to text documents and, by extension, digital text readers, such as **PDF** viewers. Renderers for **PDF** files do not usually interoperate with multimedia software, so that for example a user could not in general click on a chemical formula referenced in a research paper and then have an option of viewing a corresponding molecular-**3D** model — even if the research text and **PDB** (Protein Data Bank) or **SDF** (Structure-Data File) formats coexist in one overarching data set. The **Torq**A3R system incorporates a hybrid **PDF/SVG** format whose goal is to integrate **PDF** viewers with multimedia components. This format works somewhat by analogy to **GIS** systems: consider a **PDF** page as akin to a digital map at moderate-to-high zoom levels; then sentences are like city blocks or avenues, **PDF** page coordinates are analogous to latitude/longitude, and sentence boundaries could be highlighted via geometric formations analogous to shape files. It is easy to direct readers toward a specific page in some other publication, but it is harder to create a reference to a specific *paragraph*, or *sentence*, or even the occurrence of one specific technical term of Named Entity. The example of geotagging is useful as a point of contrast: latitude and longitude coordinates (or other **GIS** parameters, such as so-called **XYZ** triples based on a global Mercatur projection) can be specified with arbitrary precision, so it is possible with one consistent tuple of numbers to unambiguously refer to a precise point on (or at some elevation above or below) the Earth's surface. Any kind of media (text, images, videos, etc.) is able therefore establish links to **GIS** maps merely by geotagging specific entities. The **GIS** case is unusual, however, in that the entire geoinformatics field is almost entirely devoted to one specific spatial domain (viz., Earth's surface); most data spaces do not have comparable universal coordinate-systems. For example, we can cite one specific page within a text document (which is, perhaps, roughly equivalent to city or county, in the cartographic realm), but we cannot (in general) designate individual points *within* a single text page.

Via a **PDF/SVG** hybrid, however, we can cite locations internal to a document page to the scale of fractions of a typographic point, similar to latitude and longitude coordinates on a map. When displayed, **PDF/SVG** files thereby become interactive: regions such as page-areas spanned by sentences, paragraphs, or smaller text segments (such as the character-range within a technical term, keyword, acronym, or other Named Entity) can be assigned unique **SVG** identifiers and take on distinct visual representations, as a cue to the user that these specific areas have special interactive possibilities. For example, sentences might be highlighted when the mouse hovers on their interior, indicating that the sentence as a whole responds to mouse events; e.g., a context menu whose options might include copying the sentence text. This is more convenient than the current convention, where text segments are copied by laboriously positioning the cursor at the first letter in a sentence and then dragging the mouse to the end (usually just beyond a closing period or other punctuation mark). That kind of extended user action can be replaced with a single click, given proper text-encoding and cross-referencing between encoded text-segments and **PDF** page coordinates.

Text mining and open data sharing have both emerged as key facets of contemporary scientific research, for related reasons: text mining and data transparency are both techniques for expanding the scope of individual research projects, maximizing the potential for insights and discoveries made possible by pooling multiple projects' data and methodology. However, existing research-oriented technologies reveal limitations that have prevented tech mining and data transparency from benefiting science to their full potential. Key limitations include: flawed transcription of scientific text into machine-readable documents that are suitable for automated text-mining tools; lack of sufficient structure within published data to fully implement protocols related to microcitations, "nanopublications," and other strategies for designating and tracking the provenance of fine-grained data aggregates within larger-scale research environments; and the absence of rigorous protocols for integrating and cross-referencing data and text across different media and genres, including scientific literature and raw data sets along with multimedia resources (videos, **3D** models, **GIS** maps, image-series, etc.).

The limitations of existing scientific-publishing technologies, in particular, became especially evident during the **Covid19** pandemic, when scientists sought to unify researcher across many distinct disciplines — from epidemiology and virology to clinical data, and biomolecular analysis of the **SARSCOVTwo** infection mechanism. A good example of the interdisciplinary response to the early pandemic was the **CORD19** corpus ("**Covid19** Open Research Dataset"), curated by the Allen Institute for Artificial Intelligence. This project made thousands of papers about Coronavirus biology freely available, but also compiled structured versions of all articles into a machine-readable format, encouraging programmers to develop text-mining and/or **AI**-based tools for analyzing the corpus in its entirety. This project actually helped to spotlight limitations of existing publishing technology: as we reviewed in a 2022 book about **Covid19** (as well as cancer and cardiac bioinformatics[1]) the **CORD19** encoding was far from perfect, with annotation errors and inconsistencies between documents. For example, Named Entities such as chemical compounds were not correctly isolated from one document to the next, which led to missed opportunities for cross-referencing. The Allen Institute acknowledged from the outset that their text extraction capabilities were limited by the formats (primarily **PDF**) in which papers were presented, and literally issued a "call to action" requesting publishers to provide better-quality machine-readable encodings of text documents.[2] We developed a more rigorous text-encoding system for our book about **Covid19** cited above, but in general the situation as of 2023 has not changed in the intervening three years.

Anti-patterns in publishing technology going beyond text mining alone: microcitation and cross-referencing practices — especially when multiple genre and presentation/visualization media are involved — remain haphazard and often unstandardized. A comprehensive scientific and/or research corpus will likely encompass assets of diverse genres beyond just text documents: video, **3D**, **GIS**, and so forth, as well as domain-specific data sets with their own media/visualization conventions. In many cases there is no standard system for cross-referencing resources from distinct media — even when there are obvious overlaps between content expressed in one media/genre and another, such as a chemical compound named (via molecular formulae) in a scientific text and also visualized in a 3-dimensional model. Or, consider videos documenting the ecological effects of climate change, which could be cross-referenced against **GIS** maps (via latitude and longitude coordinates). In a cross-disciplinary context, many different forms of media and visuals can be relevant; **Covid19** research for instance made extensive use of **GIS** models (tracking infectivity rates and related epidemiological data across the world) and also **3D** models (of the virus's notorious "spike proteins" in particular), diagnostic images (such as "ground-glass opacity" in

[1] *See* https://shop.elsevier.com/books/innovative-data-integration-and-conceptual-space-modeling-for-covid-cancer-and-cardiac-care/neustein/978-0-323-85197-8
[2] "CORD-19: The Covid-19 Open Research Dataset" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7251955/), Section 5.2.

lung scans) and electron-microscope pictures of the virus itself. Despite the thematic correlation between these different media, however, the distinct software components which can show and analyze files and data sets of these various categories are often non-interoperable. Part of the problem is that data and file formats specific to individual media genre cannot necessarily be annotated with cross-references to other genres; but an equally significant limitation is that software is not equipped to identify and/or properly utilize such cross-references even when they are available.

The problems within text corpora and multimedia data sets are interrelated, or at least stem from similar causes: in each case there is ambiguity with respect to the individuation and description of particular information/presentation units, limiting the extent to which disparate software components can interoperate — in particular, leverage shared data elements as a bridge from one component to another. This is an underlying issue when attempting to define microcitations mutually recognized by variegated multimedia technologies. Consider again the case of geotagging video frames: restricting **GIS** coordinates to a specific geographic area is one way to isolate specific parts of a **GIS** database (or any data set with **GIS** parameters as one data attribute, which might emerge from disciplines as diverse as epidemiology, sociodemographics, environmental science, ethology, anthropology, etc.). Likewise, referring to one segment within a video represents a potential form of microcitation (as opposed to the video as a whole). A microcitation protocol in these two domains would also allow application-level cross-references, so that software components dedicated to showing maps and videos, respectively, would at least have representations of the underlying data links permitting **GIS** locale to serve as a bridge between video and geospatial media. However, *absent* a common protocol the respective software components (digital map engines and video players) cannot interoperate.

The analogous problem in the domain of text-annotations derives from inconsistent standards with respect to textual citations and references more fine-grained than traditional bibliographic and page references. As mentioned earlier, existing citation protocols do not incorporate text-spans or spatial locations more detailed than individual pages (except perhaps for specialized disciplines where natural-language texts are themselves objects of study; consider the practice of numbering individual lines of verse, within literary analysis). With the growing importance of text-mining and scientific corpus analysis, the digital granularity conventionally associated with text encoding in linguistic and humanistic contexts should be extended to academic/scientific works in general. This entails:

1. A consistent schema for encoding natural-language texts as character streams, with proper disambiguation between glyphs with similar or identical appearance but playing different discursive/semantic roles (e.g., end-of-sentence periods from abbreviation punctuation and decimal points).

2. Openly declared conventions employed by each document with respect to the hierarchical nesting between paragraphs, sentences, smaller Named-Entity-like units, and other discursive/intratextual scales. In other words, documents should clarify (as part of their encoding, markup, and/or metadata) such issues as sentence-boundaries in cases like embedded footnotes, transitions between paragraph and list-style formats, and other rhetorical devices where text deviates from the typical pattern of sentences following each other in sequence. Likewise, text content that lies outside the normal paragraph/sentence/word hierarchy (e.g., phrases appearing within figure illustrations) should be encoded with its own character-sequence contexts.

3. A "bibliospatial" coordinate system that can specify zero- and two-dimensional points/regions within the spatial extent of individual document pages to arbitrary levels of precision. For example, an instruction such as "highlight the first sentence of the second full paragraph on page 3" should be read as referencing a clearly delineated geometric extent (typically a rectilinear octagon) on one visible (canonically **PDF**) page.

4. A suite of mathematical algorithms addressing numeric properties of bibliospatial systems — e.g., second-order rectilinear Voronoi tessellation (related to the problem of mapping mouse/gesture events to the nearest bibliospatial object, such as a sentence, proximate to an event-location).

5. A full mapping between bibliospatial coordinates and encoded machine-readable text: suppose we construe text encoding in terms of character-stream sets, so that text segments are located via (what we might call) "intertextual" coordinates. Localized character streams (that could be cited, copied, cross-referenced, and so forth) could then be demarcated via start/end pairs sampled from one or more larger character streams. Interop between *machine-readable* and *human-readable* versions of any document is then necessary to fully leverage bibliospatial capabilities; in other words, a full encoding should include mappings between intratextual and bibliospatial coordinates.

6. A protocol for text-to-dataset and text-to-multimedia cross-referencing: that is, text segments marked via both bibliospatial and intratextual coordinates should be formatted in a consistent manner to be cited within computer code and multimedia assets so as to notate that a particular "subcontent," or some region smaller than a multimedia file/resource in its entirety (e.g., a stream of video frames, a set of images within a series, a location visible within one configuration of a 360° photography viewer, etc.) is thematically linked to a region of text.

When transitioning from individual documents to larger corpora or textual libraries/collections, a common encoding schema also becomes significant with respect to search queries and information retrieval. The canonical situation where these concerns come into play is that of finding documents within a collection which match a query search-term. More sophisticated search engines try to consider not only the *presence* of text snippets that match a search-term, but also the *role* of the term in a matching document — for example, a query intended to locate papers "about" **Covid19** would expect results prioritizing documents that not only contain the character-string "**Covid19**" somewhere, but also are thematically and semantically *focused on* **Covid19** (in contrast to texts that might mention the pandemic but are mostly about other subjects). Distinguishing actual semantic relevance from "false positive" matches (documents which contain a search term but without containing much information *relevant to* the search) is a separate matter from large-scale corpora queries, however, so for the current discussion we'll consider merely the problem of locating documents that do contain a particular query-term (which itself is non-trivial). The overarching problem is that — although we intuitively understand natural language as a single-dimensional flow of words and sentences — the actual data through which documents are composed, stored, and rendered (into human-readable formats, such as **PDF** pages) are more complex than simple character streams.

By way of illustration, suppose someone searches a list of publications for instances of a query term such as "Covid19" infectivity rate. The assumption behind such a search is that some documents will contain a string of characters matching the search term (perhaps allowing for small variations, such as treating "Covid" and "COVID" as the same word, or matching "infection" as well as "infectivity"). This is a reasonable assumption in the sense that as people read documents they subconsciously form a mental image of character-symbols connecting to form words, so it is usually obvious for *humans* whether a given segment of text includes a given keyword or keyphrase as a proper part. However, the ways that documents are internally stored can distort the "natural" string of characters, with markup and typographical details causing text segments to be interrupted or fragmented. The difficulties mentioned earlier vis-à-vis the **CORD19** corpus were due in part to formats such as **PDF** introducing artifacts that distort each document's underlying character sequences. A fully machine-readable text encoding would need to be a *separate* resource from human-readable documents, one which unambiguously defines a canonical character stream covering the entire document, separate and apart from markup or presentation details. For

instance, the fact that certain characters are typeset in italics or boldface should not alter how they are matched against search queries (although, in some contexts, font choices such as italicization can be relevant — suppose italics are used to state the title of a book; in this case a search might prioritize keywords that do appear within *titles*, so this particular typographic convention should be taken into account). In effect, machine-readable encoding should permit the full scope of textual content to be queried as a single character stream, while also — separate and apart from the stream itself (the separation of markup from content is often called "standoff annotation") — presenting information about text segments' typesetting and discursive details when relevant (i.e., which are book titles, Named Entities, technical terms, and other information that might come into play for search queries).

In sum, data- and resource-management for Translational Informatics naturally extends to natural language content-curation and text mining, so that software components dedicated to text documents should be intrinsic to **TSI**-related technology. As with other facets of Translational Research, computational techniques pioneered in biomedical and bioinformatics areas could certainly be extended to other scientific fields, particularly those related to community development and environmental sustainability.

Integrating Disparate Web Resources

One challenge for Sustainability technology, that can be addressed via **TSI**, is lack of interoperability between thematically or substantially connected web resources. An example from earlier — contemplating a merger of **ZOLA**, **PRISM**, and **NJ-DEP**'s land-use maps — yields a good case study: there is no conceptual reason (as opposed to fiat political boundaries) why land-use information should be split at the Hudson River. It is easy to find additional examples; consider the other relevant project with the name **PRISM**, from London, which (adapted to New York City building codes and norms) would be an excellent resource for the New York area (quite possibly, in the other direction, a version of **ZOLA** via Ember Mapbox Composer would be of value for the London conglomeration as well). Unfortunately, existing resources tend to suffer from a technological siloing effect: in this case, **ZOLA**, **NJ-DEP**, and both **PRISM**/**PRISM**s are each built with different software stacks, and there is no straightforward option to merge them, short of just porting one solution to the other's framework. There are several different steps which projects can take to mitigate this siloing effect, such as

1. Adopt standard vocabulary and conventions for **API**s and other data-sharing scenarios: Even though they offer only a restricted level of interoperability, projects overlapping with **TSI** concerns should wherever possible support **API**s (providing **API** endpoints as web addresses) and structure them around common nomenclature as recommended by organizations such as **ISI**, **WRI**, **NIEHS**, and the World Food Programme.

2. Design web sites that can readily be used as embedded components in standalone, desktop-style applications: This covers the case of application-integration achieved by accessing a select group of web applications through a self-contained desktop program rather than a generic web browser. The most common approach for web site embedding in a general-purpose, cross-platform manner is via **QtWebChannel** (mentioned earlier), which enables web pages' JavaScript assets to communicate with native-compiled code. Supporting proper embedding in this scenario involves sending JavaScript signals and giving host applications the opportunity to respond to user-events, potentially pre-empting the current web application's default handlers. When a user right-clicks on a map display, say, the host application might provide a context menu with its own action options, replacing a context menu which the web site would show ordinarily.

3. Develop reusable resources so that the capabilities exposed via web sites can be incorporated directly in standalone applications: Much of the supporting code and "business logic" behind a web service/application can often be refactored as component modules (e.g., statically linked libraries) so that multiple web sites' functionality can be aggregated into a single self-contained application. In this case, content from the individual web sites might not be read via **HTML** pages at all, but rather rendered via desktop window displays customized for users' preferences and operating system.

The second item on this list — about standalone applications hosting multiple embedded web sites — supplies a good illustration of application-level integration, as a supplement to data-level interoperablity (discussed earlier in the context of integrated event-handling). A canonical use-case is that of user actions on **GIS** map displays: almost all such events should be interpreted through the **GIS** coordinates where an action occurs, which are a product of any digital map's current display state (zoom and pan levels — the former specifying the degree of details and magnification for **GIS** elements and the latter measuring how far users have dragged the map North, South, East, or West, determining the latitude and longitude of the map's current center — plus which data layers are currently visible). Unfortunately, however, there is no standard protocol for obtaining this information once **GIS** web displays are embedded in desktop software. The host application knows window coordinates where events take place, but there is no single (or guaranteed) algorithm to convert those coordinates to **GIS** locations (or street addresses). Even where a host application can infer this data indirectly from an embedded web site (e.g., via **URL** patterns) the techniques for doing so will typically be idiosyncratic for each embedded resource, so each individual web site requires manual coding for the relevant event-handlers.

Adopting a common embedding protocol, by contrast, would enable embedded web sites to notify host applications of required event meta-data via a shared model so that new peer web sites could be added with minimal extra programming. With such techniques, geographically related sites could be stitched together almost seamlessly, so one could for instance transition from land-use queries on New York territory to New Jersey via a front-end interface (in terms of context menus, secondary windows, data-layer styling, etc.) which is largely the same on either side of the Hudson River. Such a common interface style could then be extended beyond digital maps proper to other assets that may be used in conjunction with **GIS** services — for example, linking place-names as **GIS** locations (with implicit latitude and longitude points/boundaries) to text-segments in **HTML** or **PDF** files (**GIS** elements, such as names of counties and cities or towns/boroughs, becoming treated as Named Entities with a specific kind of associated metadata, viz., geographic coordinates and boundary-shapes).

Whereas syncing information sources at the data-integration level only involves modifying or extending a resources' public interface (**API** endpoints and **URL**s, for example), supporting more detailed application-integration along the lines just described can require more low-level programming within web resource's implementations. We propose to provide code libraries to facilitate such **TSI** extensions as part of our overall data-integration project.

Government information services such as we proposed earlier can also serve as Reference Implementations demonstrating **TSI** protocols along these lines. This would be an added benefit to a services curated under the aegis of Bergen County and or the state of New Jersey, insofar as the technology employed could become a template for use by other jurisdictions, including those outside the New York metro areas entirely. **TSI**-focused components engineered in New Jersey may likewise by unified with Ember Mapbox Composer, from NYCPlanning, for a multi-purpose service addressing both zoning/land-use and environmental/sustainability concerns.